FINAL REPORT APPENDICES
FOR
SILICON-BASED QUANTUM MOS TECHNOLOGY DEVELOPMENT

CONTRACT NO. F49620-96-C-0006
13 December 1995 - 31 December 1999

Prepared For
Department of the Air Force
Air Force Office of Scientific Research (AFOSR)
Bolling Air Force Base, DC 20332-0001

Prepared by:
Raytheon Systems Company
P.O. Box 660246, M.S. 35
Dallas, TX 75266

Principal Investigator
Dr. Alan Seabaugh
Present address University of Notre Dame

20000407 044

Subcontracts:

Texas A&M University	Prof. Wiley Kirk
University of Delaware	Prof. James Kolodzey
University of Illinois	Prof. Joseph Lyding
University of Michigan	Prof. Pinaki Mazumder
University of Texas at Dallas	Prof. William Frensley

Distribution Statement A. Approved for public release; distribution is unlimited

DTIC QUALITY INSPECTED 1

AFRL-SR-BL-TR-00-

. 0091

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruct
data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspe
this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188),
4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to c
valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* 07 March 2000	**2. REPORT TYPE** Final Report	**3. DATES COVERED** *(From - To)* 13 Dec. 1995 - 31 Dec. 1999

4. TITLE AND SUBTITLE
Silicon-Based Quantum MOS Technology Development

5a. CONTRACT NUMBER
F49620-96-C-0006

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Dr. Alan Seabaugh, Raytheon Systems Company, Prof. Wiley Kirk, Texas A&M University, Prof. James Kolodzey, University of Delware, Prof. Joseph Lyding, University of Illinois, Prof. Pinaki Mazumder, University of Michigan, Prof. William Frensley, University of Texas at Dallas

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Raytheon Systems Company
13532 N. Central Exp. M/S 35
Dallas, Texas 75243

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force
Office of Scientific Research
801 N. Randolph St. Room 732
Arlington, VA. 22203-1977

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release: Distribution is unlimted

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This program demonstrated the first compatible resonant tunneling diodes (RTD) for integration with CMOS (complementary metal oxide semiconductor) technology and designed circuits to exploit and evaluate the speed-power advantage of integrated RTD and CMOS devices. The first Si-based resonant interband tunnel diodes were demonstrated with peak current density greater than 10^4 A/cm^2; peak-to-valley current ratio exceeding 2 was achieved at room temperature. The $SiO_2/Si/SiO_2$ RTD was studied extensively in a variety of forms, however no room temperature negative differential resistance was observed in the $SiO_2/Si/SiO_2$ system. Other double barrier systems were also studied including ZnS/Si, CaF_2/Si, and Si_3N_4/Si without evidence of resonant tunneling. Circuit designs using SPICE showed that integration of tunnel diodes with CMOS provides typically more than 2x speed and density improvement for analog and digital circuits and more than 20x reduction in static power dissipation for memory over conventional circuit approaches. A roadmap was developed outlining the steps to commercialization.

15. SUBJECT TERMS
Resonant Tunneling, CMOS, QMOS, Nanoelectronics, Integrated Circuit

16. SECURITY CLASSIFICATION OF: **Unclassified**			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr. Gerald Witt
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)* 703-696-8571

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

APPENDICES

Appendix 1

BEYOND-THE-ROADMAP TECHNOLOGY: SILICON HETEROJUNCTIONS, OPTOELECTRONICS, AND QUANTUM DEVICES

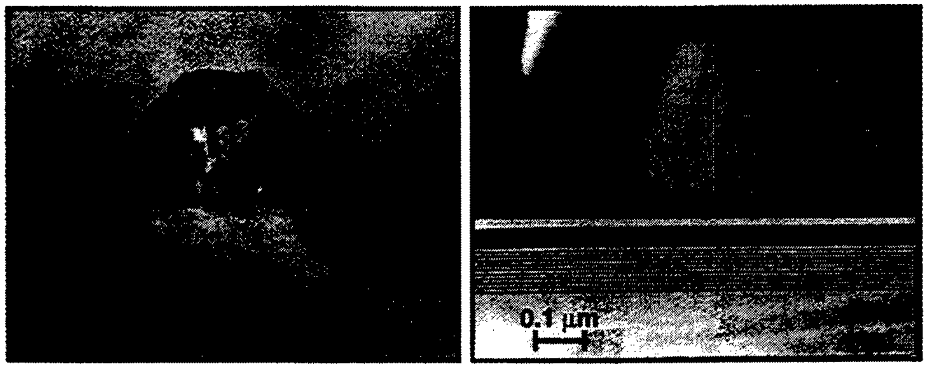
Alan Seabaugh, Roger Lake, Bobby Brar, Robert Wallace†, and Glen Wilk†
Raytheon TI Systems, Dallas TX 75243, seabaugh@resbld.csc.ti.com
†Texas Instruments Incorporated, Dallas TX 75243

ABSTRACT

The roadmap for silicon device technology has been drawn, extending to the year 2010, and featuring a CMOS transistor with a gate length of 0.07 µm [1]. Beyond this point, silicon heterojunctions could provide a means to further device scaling. Silicon heterojunctions could also bring new devices to the silicon substrate including light emitters and detectors, and resonant tunneling diodes (RTDs). Today SiGe/Si and SiGeC/Si heterojunctions are receiving the greatest attention, but heterojunctions now being developed to realize silicon RTDs are increasing the heterojunction options for silicon-based quantum-well and optical devices. Here we outline the fundamental device requirements for silicon optical and tunneling devices and describe progress on silicon heterojunction development towards demonstration of silicon-based RTDs. Materials now under study include, ZnS, crystalline oxides and nitrides; new processes could provide methods for forming crystalline materials over amorphous barriers.

INTRODUCTION

Striking differences between high performance silicon and compound semiconductor device approaches can be seen in the cross-sectional micrographs of Figure 1. Shown at left is an 0.18 µm MOSFET (metal-oxide-semiconductor field-effect transistor) with a polysilicon gate and a gate oxide of approximately 5 nm. In the MOSFET, the energy band profiles in both the lateral and vertical directions are engineered by close control of the dopant profiles. The dark contrast in the transmission electron micrograph (TEM) of Figure 1(a) is caused by strain induced by the



7a1125.ppt

(a) (b)

Figure 1. Comparison of (a) the silicon MOSFET and (b) the InP-based HFET at the same scale and contrasting the differences between dopant and bandgap engineering.

Mat. Res. Soc. Symp. Proc. Vol. 486 © 1998 Materials Research Society

heavy source/drain implants. In contrast, Figure 1(b), an InP-based 0.2 μm gate heterojunction field-effect transistor (HFET) is shown at the same scale through a composite figure consisting of a scanning electron micrograph (SEM) of the gate above a TEM of the underlying epitaxy appropriately scaled. A silicon heterojunction technology which could harness both dopants and heterojunctions would enable a new flexibility in the optimization and continued scaling of MOS device technology.

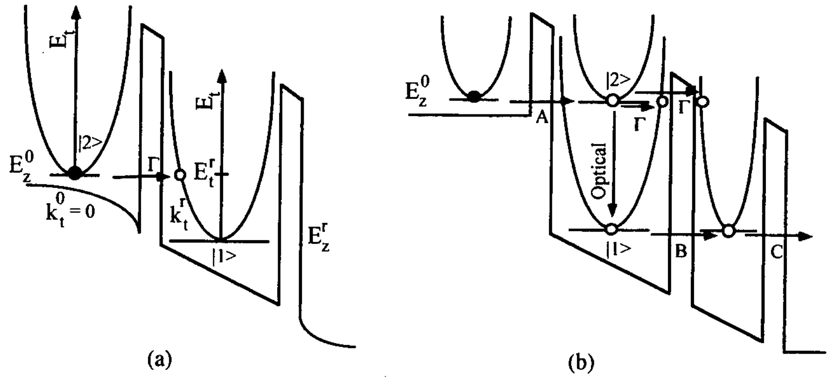
Heterojunctions on silicon could also enable the construction of tunable silicon light emitters, detectors, optical interconnects, and tunneling devices. High performance resonant and interband tunneling diodes, when combined with any transistor technology allow circuit compaction which can provide a speed increase of 2 to 5× and a reduction in power dissipation by a similar factor. Circuit simulations and layouts of RTDs in a CMOS process have shown that static power dissipation can be reduced by approximately 100× in embedded memory/logic applications. For more information on resonant tunneling circuit technology consult references [2-4].

In this paper, we outline the materials requirements for silicon based heterojunction device development and review current progress in the development of silicon heterojunctions for RTDs. We find that the fundamental materials requirements for Si tunneling devices are the same as for photonic sources; this is easy to accept since the RTD is the first step in the fabrication of the silicon-based superlattices. In the experimental section, we outline recent progress in the development of amorphous, crystalline, and mixed amorphous/crystalline heterojunctions toward the demonstration of resonant tunneling diodes.

Relationship Between Optical and Tunneling Devices

For devices which are not small in cross section (greater than ~100 nm^2), there are two classes into which we can group both tunnel devices and optical devices such as sources and detectors. The two classes are (i) unipolar or intraband and (ii) bipolar or interband. The energy and momentum conservation laws which govern the performance of both tunnel devices and optical devices are the same for each class. For intraband devices in which the transport or absorption and emission takes place solely within either the conduction or valence band, both conservation of total energy and transverse momentum are required to obtain good performance. This is because in intraband devices, there are no energy gaps, only momentum gaps. For interband devices in which electrons tunnel or optically recombine across the bandgap, only conservation of total energy is required to obtain good performance. Thus, interband devices, either electrical or optical, tend to be more resilient to the effects of disorder since they require one less conservation law to work.

Figure 2(a) illustrates the energy and momentum conservation requirements of a unipolar RTD. The horizontal line in the well between the two barriers is generally referred to as the energy level of the resonant state in the well or the "resonant energy." However, it is actually the bottom of a 2-dimensional (2D) subband of states represented by a parabola. In Figure 2(a), the RTD is biased in the valley current region such that the electron in the emitter with zero transverse momentum is incident at an energy above the "resonant energy." In an ideal epitaxial structure, the periodic crystal potential is undisturbed in the transverse plane, therefore transverse crystal momentum is conserved, and the process illustrated in Figure 2(a) is not allowed. In the presence of disorder from, for example, impurities, alloy, or interface roughness, the incident electron can pick up a Fourier component of the disorder potential and scatter into a transverse momentum state k_t^r. This is the process by which elastic scattering from disorder contributes to the valley current.



(a) (b)

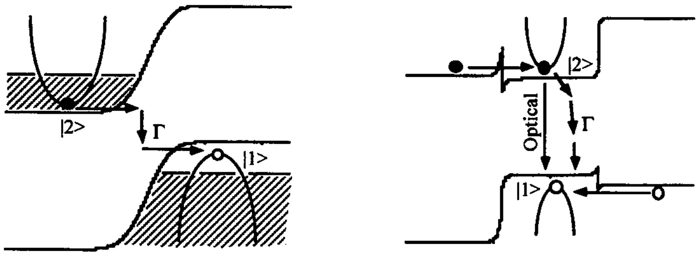
Figure 2. Comparison of unipolar intraband tunnel and optical devices: (a) resonant tunneling diode and (b) optical light emitter. The parabolas shown on the energy band diagrams represent the transverse kinetic energy. An off-resonant incident electron in (a) with longitudinal energy E_z^o and zero transverse kinetic energy cannot tunnel and conserve total energy and transverse kinetic energy. The processes labeled Γ can occur if transverse momentum is not conserved such as in amorphous materials; breaking transverse momentum conservation results in an increase in valley current in RTDs and a reduction in the quantum efficiency of optical devices.

Therefore, any process, either elastic or inelastic, that destroys transverse momentum conservation, reduces the peak-to-valley ratio (PVR) of a unipolar RTD.

In a similar way, lack of transverse momentum conservation reduces the performance of intraband optical devices which are based on electron transitions from one "resonant state" to another. A schematic example of an infrared source is shown in Figure 2(b). An electron is injected from the emitter into the second resonant state in the left well. Ideally, it will then optically recombine with the hole in the first resonant state and then tunnel out into the collector following the path A-Optical-B-C. However, if transverse momentum conservation is broken, the processes labeled Γ can also occur. These processes reduce the quantum efficiency of the device.

For interband tunnel diodes such as the Esaki diode [5], the quantum well Esaki diode [6], or the resonant interband tunnel diode [7], only conservation of total energy is required to obtain good performance since a true energy gap exists between the electron and hole states. Figure 3 compares the (a) Esaki diode biased into the valley current region with the (b) interband optical source. Neither device requires transverse momentum conservation to work. The tunneling event labeled Γ in Figure 3(a) is prohibited in an ideal Esaki diode by energy conservation. Only processes that destroy the energy conservation of the tunneling electron such as phonon or photon emission reduce the peak-to-valley current ratio and device performance. Therefore, interband tunnel devices should be more resilient to scattering due to disorder than intraband tunnel devices. This is evident for an Esaki diode since the active region is doped to concentrations of 10^{20} cm^{-3}; such a high doping in the well of an RTD would destroy its NDR.

In the same way, interband optical devices rely only on energy conservation for their operation. Figure 3(b) shows an interband optical source. Electrons are injected from the *n*-contact on the left into the well where they optically recombine with holes injected from the *p*-contact on the right. The quantum efficiency of this device is not effected if electrons do not conserve their transverse momentum. However, any nonradiative recombination reduces the efficiency of this device;



(a) (b)

Figure 3. Comparison of bipolar interband tunnel and optical devices: (a) Esaki diode biased into the valley current region and (b) optical light emitter. The Esaki diode and interband optical source do not require transverse momentum conservation.

nonradiative recombination requires an inelastic process for an electron to lose the energy of the bandgap.

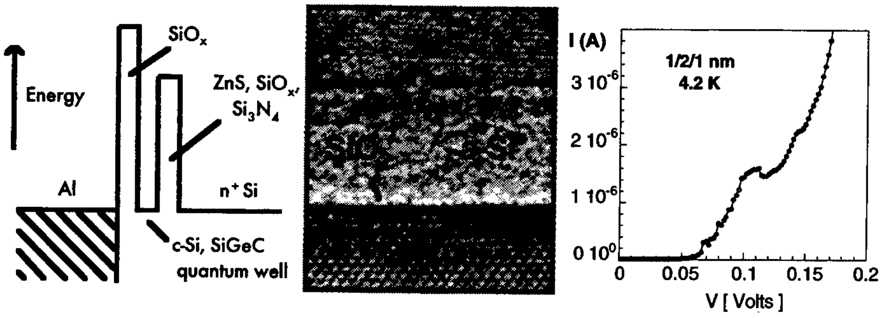
The performance of both tunnel devices and optical devices is limited by the nonradiative recombination from state |2> in subband 2 to state |1> in subband 1. This recombination process is labeled as the Γ process in Figures 2 and 3. The nonradiative recombination rate from subband 2 to 1 in intraband devices, Figure 2, is fast since the subbands are degenerate. The nonradiative lifetime resulting from phonon emission is on the order of 1 ps. For interband devices, the nonradiative lifetime is much longer. For an interband device, an electron must emit multiple phonons to lose an energy on the order of the bandgap and the recombination process is mediated by localized states in the bandgap. Nonradiative lifetimes of interband devices can be on the order of 1 ns resulting in a better quantum efficiency than intraband devices. The key point is that the performance of interband devices is not effected by momentum conservation, therefore interband devices should be more resilient to the presence of disorder.

EXPERIMENT

Amorphous Si/SiO_2 Superlattices and Resonant Tunneling Diodes

Experiment is consistent with the above theoretical discussion for both amorphous Si/SiO_2 superlattices and resonant tunneling diodes. There have been a number of studies of the optical properties of Si/SiO_2 superlattices to look for quantization and resonance effects [8,9]. Core level x-ray spectroscopy and photoluminescence gives strong evidence that quantization does occur, and that it follows the common inverse-square relationship to the well width [8]. The observation of state-quantization in these superlattices is to be expected despite the fact that the Si/SiO_2 superlattices are amorphous since the optical processes require only conservation of total energy.

In the case of resonant tunneling diodes formed out of the amorphous SiO_2/Si/SiO_2 double barrier there have been no high-PVR demonstrations of resonant tunneling and negative differential resistance (NDR). This is also consistent with theoretical expectations, since the observation of NDR requires conservation of both total energy and transverse momentum and transverse momentum is not conserved in amorphous materials. The random potential resulting from the amorphous Si/SiO_2 heterostructure breaks the translational periodicity and thus the transverse momentum conservation required for high PVR in RTDs. Typical findings for this system are illustrated in Figure 4.



(a) (b) (c)

Figure 4 Silicon resonant tunneling double barrier: (a) schematic energy band diagram, (b) transmission electron micrograph of an amorphous resonant tunneling heterostructure with oxide thicknesses of 0.9 nm and quantum well width of 2 nm, and (c) measured current-voltage characteristics of an amorphous RTD like the device pictured in (b) at low temperature showing negative differential resistance (150 μm device diameter). The peak-to-valley current ratio is low and the characteristic varies widely from device to device; The peak voltage does not scale inversely as the square of the quantum well width. The negative differential resistance at 1.1 V is probably associated with zero-dimensional states (due to dangling bonds, interface or impurity states, or defects) located within the double barrier.

The amorphous SiO$_2$ barriers shown in Figure 4(b) were grown by a room temperature ultraviolet (UV) ozone process, but similar results are also obtained for oxides grown by thermal dry oxidation. The amorphous Si quantum well was deposited at room temperature in an MBE (molecular beam epitaxy) chamber. The aluminum top contact which serves as the positive terminal or collector of the RTD is typically formed *in situ* before removal from an ultrahigh vacuum system which allows oxidation, MBE, and metallization processes without breaking vacuum.

Numerical simulations of the current-voltage (I-V) characteristics for amorphous devices has shed some light on the experimental RTD results [10]. Current-voltage calculations that include scattering in the amorphous material as a random potential show clearly that the resonance width is directly related to the strength of the scattering as shown in Figure 5. Carrier mobility in the material is shown to be a useful parameter for characterizing the scattering in the amorphous structure. *The essence of the theoretical analysis is that crystalline material, at least in the well region, is necessary to observe resonant tunneling in these materials.* Simply speaking the lack of crystalline periodicity results in increased carrier scattering thereby causing the resonances in the structure to broaden. The broad resonances manifest themselves in poor or non-existent peak-to-valley ratios for resonant-tunneling diodes. However, one important additional conclusion from the theoretical analysis is that scattering sites in the *barriers* do not significantly effect the electron transport because of the (exponentially) reduced amplitude of the electron wavefunction within the barrier. This has motivated the fabrication approach which follows.

Formation of Crystalline Silicon on Ultrathin Voided Oxides

The barriers used in a resonant tunneling diode or superlattice can in principle be porous i.e. the barrier such as SiO$_2$ can have nanometer sized voids through which a silicon overlayer can be

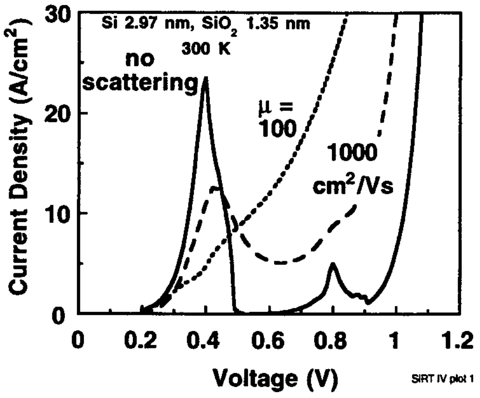


Figure 5. Computed current-density voltage charactistic showing the improvement of peak-to-valley current ratio with electron mobility (crystallinity) in the quantum well where a random potential proportional to the carrier mobility is used to explore the effects of scattering in the amorphous heterostructure [10].

nucleated. If the voids in the tunneling barrier have openings of size smaller than the electron wave packet spread, the tunneling barrier height is not reduced and the nucleated Si quantum is crystallographically aligned to the underlying silicon. A method for forming these voids in ultrathin layers has been developed [11,12] using the thermal desorption of SiO at elevated temperatures in ultrahigh vacuum to form the voids, see Figure 6. In practice, as detailed in [11,12], we have found that there is a direct relationship between the size of the voids and their density, and that crystalline Si overgrowth on Si (100) is only obtained for large void densities (exceeding approximately 2500/μm^2) and large void area (exceeding approximately 400 nm^2). This void area is approximately one hundred times too large for practical growth of voided-barrier resonant tunneling diodes and optical sources.

Since the work of references [11] and [12] void formation and growth experiments in ultrathin thermal oxides on the Si (111) surface have been conducted and shown to exhibit similar behavior to that observed on Si (100). These results are summarized in Figure 7 where an increase in void

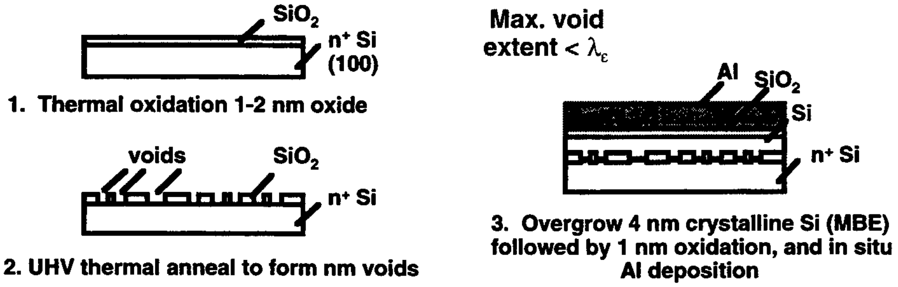

SiO2
n+ Si (100)
1. Thermal oxidation 1-2 nm oxide
voids
SiO2
n+ Si
2. UHV thermal anneal to form nm voids
Max. void extent < λε
Al
SiO2
Si
n+ Si
3. Overgrow 4 nm crystalline Si (MBE) followed by 1 nm oxidation, and in situ Al deposition

Figure 6. Schematic cross section of a process by which crystalline silicon is nucleated through a voided amorphous barrier.

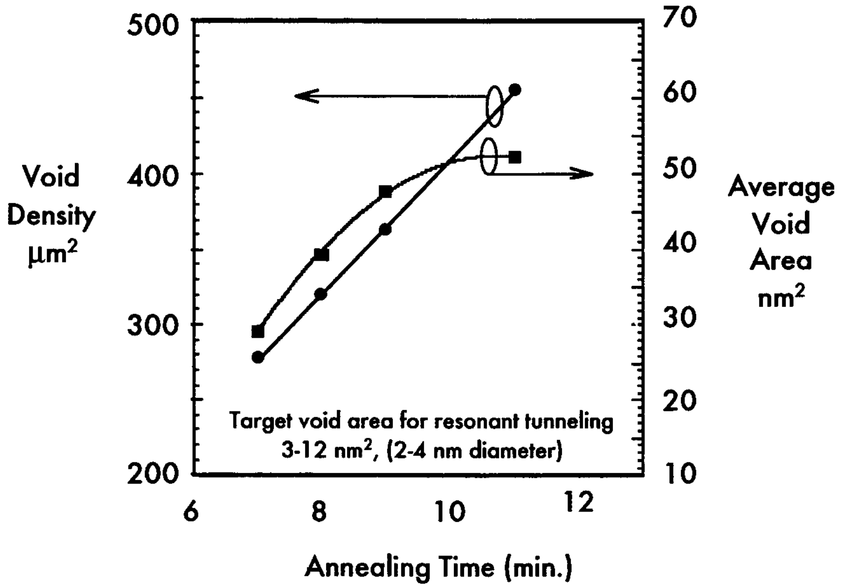
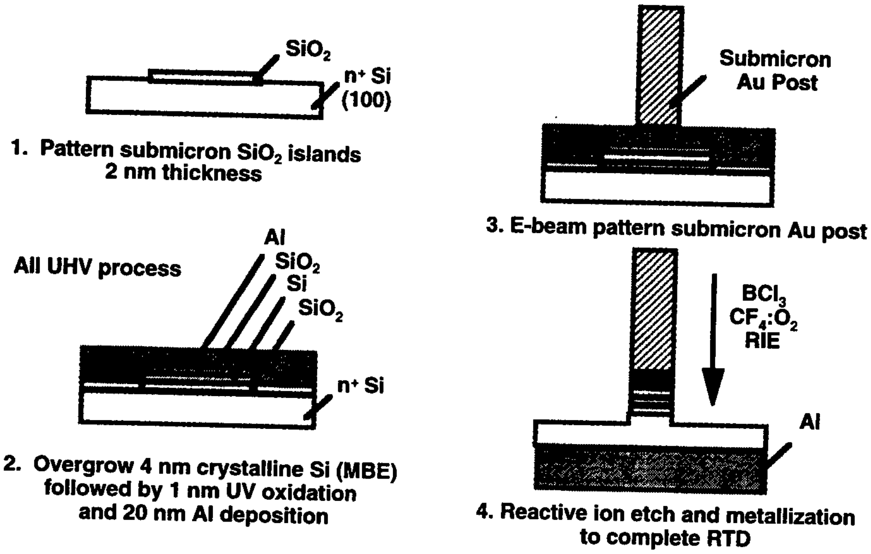


Figure 7. Dependence of the density and area of voids formed in a 1 nm SiO_2 film on (111) Si on annealing in vacuum at 750 °C for the indicated times. Void area and density is determined by scanning electron microscopy.

density and size are again observed with anneal time. Interestingly, the formation of observable voids requires a longer period of time at temperature, suggesting that the void initiation on the (111) surface is more difficult than on the (100) surface. Like the (100) surface, the smallest oxide void areas found on the (111) surface (~29 nm^2) do not meet the targets (3-12 nm^2) for use in oxide/silicon/oxide RTDs.

Formation of Crystalline Silicon on Amorphous Barriers by Lateral Overgrowth on SiO_2 Islands

A second method now under investigation for resonant tunneling diodes using $SiO_2/Si/SiO_2$ double barriers is illustrated in Figure 8. In this case, the crystalline Si quantum well is formed over lithographically-defined oxide islands, by molecular beam lateral epitaxial overgrowth. The greater silicon surface provides the seed for lateral crystalline overgrowth of the oxide islands. Since the silicon overgrowth process is formed on a micron-scale to submicron scale and the RTD emitter contacts are as small as 0.01 square microns, the current through a single RTD can easily become less than 10 fA if the oxide thicknesses become large (>2 nm). For this reason arrays of RTDs are formed in addition to the single device: array sizes of 10, 100, 1000, and 10,000 RTDs allow parallel measurements of the average current through an RTD when the single device current is too low to be easily measured individually. The most critical steps are those which lead to step 1. To begin, a thermal oxide of about 4 nm is grown on an HF terminated Si (100) surface. Submicron islands are then formed by photolithography and the oxide mesas are formed by wet etching in buffered HF. Just prior to loading for growth of the quantum well, the 4 nm oxide is


SiO$_2$
n+ Si
(100)
Submicron
Au Post
Al
All UHV process
SiO$_2$
Si
SiO$_2$
n+ Si
BCl$_3$
CF$_4$:O$_2$
RIE
Al

1. Pattern submicron SiO$_2$ islands 2 nm thickness

2. Overgrow 4 nm crystalline Si (MBE) followed by 1 nm UV oxidation and 20 nm Al deposition

3. E-beam pattern submicron Au post

4. Reactive ion etch and metallization to complete RTD

Figure 8. Schematic diagram of a process under development for realization of resonant tunneling diodes with amorphous Si barriers and crystalline Si quantum well.

thinned by wet chemical etching again in buffered HF to thin the oxide below 2 nm and hydrogen terminate the greater Si surface for growth. Scanning electron micrographs of the overgrown oxide/silicon/oxide double barrier are shown in Figure 9. This image is taken after the emitters have been deposited and before reactive ion etching to form the active device areas; it is also taken

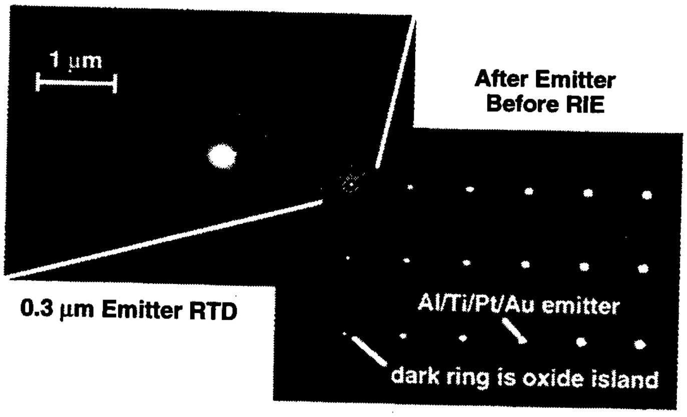

1 µm
After Emitter
Before RIE
0.3 µm Emitter RTD
Al/Ti/Pt/Au emitter
dark ring is oxide island

Figure 9. Scanning electron micrographs of the SiO$_2$/Si/SiO$_2$ resonant tunneling diode (wafer A4) taken after emitter formation over the double barrier. The first oxide islands are seen as dark regions about the submicron emitters.

from a region of the wafer from which the *in situ* deposited aluminum has been masked away. What can be seen in the lower right micrograph are 18 different RTD emitters, the first row with 0 degree orientation, the second row with a 15 degree rotation, and the third row with a 30 degree rotation. About each emitter a dark ring is seen corresponding to the location of the first and underlying oxide island. The expanded region in the upper left was taken first and the charging of the second surface oxide is apparent in the second, lower right figure where the charged surface has more emission than the previously uncharged surface. This is qualitative confirmation of the presence of the second oxide.

Crystalline Silicon Heterojunctions

Silicon heterojunction research has recently focused on the SiGe/Si and SiGeC/Si systems which should provide conduction band offsets as high as approximately 0.5 eV in the limit of high Ge content where only thin layers can be grown without lattice relaxation. Silicon germanium and its alloys with carbon [13] do not provide the broad bandgap flexibility enjoyed by the compound semiconductor systems. The SiO_2/Si heterostructure with a band offset of 3.2 eV provides an excellent barrier for use as a gate oxide and when formed at thickness below approximately 4 nm provides textbook direct tunneling barrier [14], but bandgap engineering are difficult due to it's lack of crystallinity and the difficulty with silicon overgrowth.

A single crystal silicon oxide would provide several opportunities for advanced device development. First and foremost, given a crystalline silicon oxide barrier, it should then be possible to grow single crystal Si over the insulating oxide. We have developed a method for growing a single crystal silicon oxide layer directly on a Si (100) substrate, which is not fully SiO_2, but rather is a suboxide SiO_x ($0 < x < 2$) which is still insulating and could serve as the insulating barrier while still provide an epitaxial template for subsequent overgrowth of a crystalline Si quantum well. Crystalline SiO_x films (200 nm thick) have now been demonstrated, with oxygen contents ranging from 20 - 40%, as determined by Rutherford backscattering (RBS) measurements. The crystal quality of the top Si layer has also been measured by RBS, and although it is defective (several films had $c_{min} = 20\%$, where $c_{min} = 3\%$ for a perfect epitaxial Si film). Single-barrier tunnel diodes have been constructed and show that the desired insulating property of the barrier is maintained with a barrier height of approximately 0.5 eV.

The need for crystalline insulating layers on Si which can be used as seeds for crystalline Si overgrowth has motivated us to consider the growth of crystalline silicon nitride. We have now demonstrated growth of an ultrathin crystalline nitride on a Si (111) substrate. The crystalline layer was grown by thermal nitridation in an NH_3 ambient, at a partial pressure of 1×10^{-5} Torr. In order to obtain growth of a crystalline nitride, the substrate must be heated above the 1 x 1 to 7 x 7 transition on Si (111), which occurs at 840 °C. Above this temperature, an 8/3 x 8/3 reconstruction is obtained, as shown in the 36 nm x 36 nm scanning tunneling micrograph in Figure 10. This crystalline insulator could provide an epitaxial template for subsequent overgrowth of a crystalline Si quantum well. Although the lattice mismatch between Si_3N_4 and Si is large (~20%), this reconstruction corresponds to a stoichiometry of Si_3N_5.

The crystalline oxides and nitrides offer one path to forming an epitaxial template for overgrowth of crystalline silicon. With an energy gap of 3.6 eV and a lattice constant of 5.42 Å, zinc sulfide (ZnS) is an excellent candidate for use as an insulating barrier that is lattice-matched to silicon. Recently, high quality ZnS layers on silicon have been realized by initiating MBE growth on a vicinal Si (001) surface (4° off-cut towards the $[0\bar{1}1]$direction) that has been terminated with

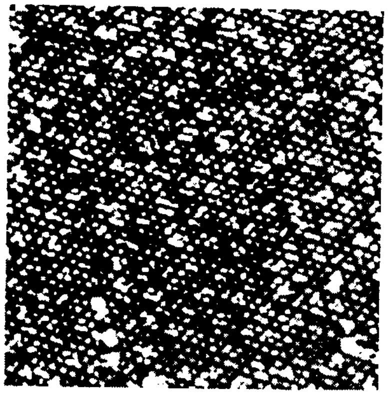


Figure 10. Scanning tunneling micrograph showing an 8/3 x 8/3 reconstruction of an epitaxial silicon nitride film grown on a Si (111) substrate.

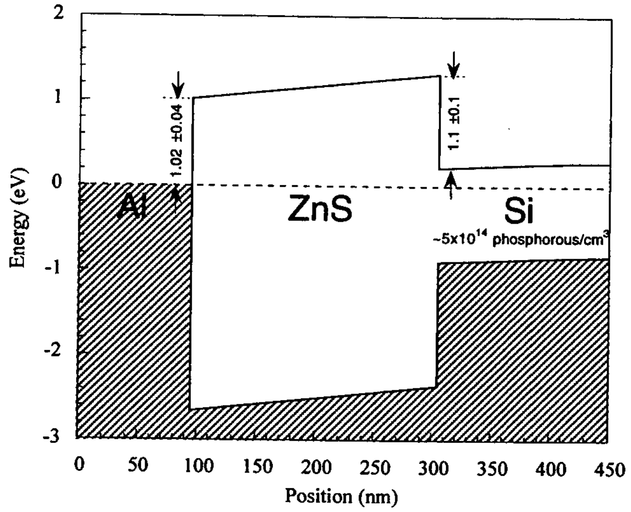

2
1
0
-1
-2
-3
Energy (eV)
1.02 ±0.04
1.1 ±0.1
Al
ZnS
Si
~5x10^{14} phosphorous/cm^3
0
50
100
150
200
250
300
350
400
450
Position (nm)

Figure 11. Zero-bias energy band diagram for Al/ZnS/As(1ML)/Si structure.

a single monolayer (ML) of As [15,16]. The band offsets for this heterojunction have also been determined [17]. Figure 11 shows the energy band diagram for the Al/ZnS/As(1ML)/n-Si heterostructure. The approximately 1 eV ZnS barrier heights to Al and Si, being much larger than kT at room temperature, are encouraging for the future room-temperature operation of ZnS/Si RTDs and superlattices.

CONCLUSIONS

We have shown that the energy conservation requirements for tunneling and superlattice devices are the same and that amorphous barriers should be acceptable for superlattice optoelectronic and resonant tunneling devices, so long as crystalline Si overlayers can be overgrown with crystalline registry to the underlying silicon substrate. Towards this end we have described two processes now under development to form resonant tunneling diodes using amorphous silicon dioxide tunnel barriers.

Silicon heterojunction growth is still in its early stages of exploration. Among the most promising new junction materials for Si are crystalline oxides, crystalline nitrides, and II-VI compounds such as ZnS. Other materials such as aluminum nitride, cerium oxide, and gallium phosphide deserve attention.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of AFOSR and DARPA under contracts F49620-95-C-0001 and F49620-96-C-0006. Special thanks for technical support are expressed to M. Coviello, M. Howell and P. Stickney, and to a large group of people who have shaped the ideas and contributed to the effort, including B. Bate, T. Blake, T. Broekaert, G. Frazier, W. Kirk, R. Steinhoff, H.-L. Tsai, P. van der Wagt, and Y. Wei.

REFERENCES

1. National Technology Roadmap for Semiconductors, SIA (1994).
2. A. C. Seabaugh, B. Brar, T. Broekaert, G. Frazier, and P. van der Wagt, 1997 GaAs IC Symp., pp. 119-122.
3. T. P. E. Broekaert, B. Brar, P. van der Wagt, F. Morris, T. Moise, G. Frazier, and E. Beam III, 1997 GaAs IC Symp., pp. 187-190.
4. P. van der Wagt, A. Seabaugh, and E. Beam, III, 1996 IEDM Tech. Dig., pp. 425-428.
5. L. Esaki, Phys. Rev. **109**, 603-604 (1958).
6. H. H. Tsai, Y. K. Su, H. H. Lin, R. L. Wang, and L. Lee, IEEE Electron Device Lett. **15**, 357-359 (1994).
7. J. R. Söderström, D. H. Chow, and T. C. McGill, Appl. Phys. Lett. **55**, 1094-1096 (1989).
8. Z. Lu, D. J. Lockwood, and J. -M. Baribeau, Nature **378**, 258-260 (1995).
9. M. Hirose, M. Morita, and Y. Osake, Jpn. J. Appl. Phys. **16** Suppl. 16-1, 561-4 (1977).
10. R. Lake, B. Brar, G. D. Wilk, A. C. Seabaugh, and G. Klimeck, 1997 Int. Symp. Comp. Semicond.
11. Y. Wei, R. M. Wallace, and A. Seabaugh, Appl. Phys. Lett. **69**, 1270-1272 (1996).
12. Y. Wei, R. M. Wallace, and A. C. Seabaugh, J. Appl. Phys. **81**, 6415-6424 (1997).
13. J. Kolodzey, F. Chen, B. A. Ormer, D. Guerin, and S. I. Shah, Thin Solid Films **302**, 201-203 (1997).
14. B. Brar, G. D. Wilk, A. C. Seabaugh, Appl. Phys. Lett. **69**, 2728-2730 (1996).
15. L. T. Romano, R. D. Bringans, X. Zhou, and W. P. Kirk, Phys. Rev. B **52**, 11202- (1995).
16. X. Zhou, S. Jiang, and W. P. Kirk, J. Appl. Phys. **82**, 2251-2262 (1997).
17. B. Brar, R. Steinhoff, A. Seabaugh, X. Zhou, S. Jiang, and W. P. Kirk, 1997 Int. Symp. Comp. Semicond.

Electrical Characteristics of High-Quality Sub-25-Å Oxides Grown by Ultraviolet Ozone Exposure at Low Temperature

G. D. Wilk and B. Brar

Abstract— **We have developed a method for controllably and reproducibly growing self-limiting ultrathin oxides with excellent electrical properties in the range ~10–25 Å thick at temperatures ranging from 25 to 600 °C, respectively, using an ultraviolet ozone (UVO_3) oxidation process. The self-limiting thickness depends primarily on the substrate temperature, allowing ultrathin oxide growth with precision and reproducibility using this UVO_3 process. Oxides grown by this method are comparable in electrical quality to thermal oxides, with similar leakage current densities and breakdown fields $E_{BD} > 10$ MV/cm. Current–voltage ($I-V$) analysis shows oxide thickness uniformity to within 1% from center to edge of a 4-in wafer. Capacitance–voltage ($C-V$) characterization of ~25 Å oxides shows excellent saturation behavior, with low midgap interface trap densities and no hysteresis or dispersion.**

IT HAS BEEN shown that a 15 Å oxide gate dielectric can be used for transistor operation, despite a high gate leakage current [1]. Investigations into transistors with such thin gate oxides has been necessitated by the roadmap for CMOS scaling into the regime of sub-0.1-μm device feature sizes. These scaling demands have initiated much research in the area of ultrathin oxides and their properties [2], [3]. Although oxides below 15 Å thick result in unacceptably high leakage currents for gate insulator purposes, an oxide layer of this thickness may be very useful as the underlayer of a two-layer dielectric stack, such as in oxide-nitride structures. In this way, the excellent Si-SiO_2 interface properties are maintained while a higher dielectric constant material is deposited on top of it [4]. Regarding reliability, concerns over pinhole formation during oxidation have been raised for ultrathin thermal oxides [5]. In addition, the Si-SiO_2 interface roughness will play an increasingly important role in sub-0.1-μm device performance, and alternative pre-cleaning procedures to standard HF-etching have been studied in which atomically flat surfaces can be achieved at relatively low temperatures [6], [7].

Device-quality gate oxides in the 10–25 Å range require extremely precise control of the oxidation process, as a variation of only a few percent in oxide thickness may be required. In this regime, even small thickness variations within a wafer or from wafer to wafer are significant, because the direct tunneling current increases by an order of magnitude for every ~2 Å decrease in oxide thickness [8]. A major difficulty in obtaining reliable gate oxides in the 10–25 Å range is the growth process. Standard techniques such as furnace and rapid thermal oxidation will have difficulty producing these oxides due to slow ramp-up times and wafer-to-wafer growth nonuniformities, respectively.

The UVO_3 process has a potentially large advantage over traditional thermal oxidation techniques because it is inherently self-limiting, and will therefore yield more uniform oxide thicknesses. Previous work [9]–[12] on oxides grown by the UVO_3 process concentrated on thicker oxides grown at generally higher temperatures, with little or no electrical characterization, and there was no investigation into the uniformity and thickness accuracy attainable by this process. Oxide reliability [11] was shown to improve with UVO_3 exposure compared to dry O_2 oxidation.

For scaled CMOS applications, a low-temperature, self-limiting oxidation process in the 10–25 Å range is desirable. In this Letter, we demonstrate a UVO_3 oxidation process to grow uniform, high-quality ultrathin oxides in the range 10–25 Å, with substrate temperatures below 600 °C. The oxides were evaluated by electrical measurements, transmission electron microscopy (TEM) and spectroscopic ellipsometry.

The oxides were grown on nominally flat, n-type Si(100) wafers with $\rho = 0.01$–$0.02\ \Omega$ cm. The as-received wafers were prepared by a buffered 2% HF etch to remove the native oxide and provide an H-terminated surface. The measured rms roughness is ~1 Å across the wafers. The wafers were transferred into a high vacuum chamber with a base pressure below 5×10^{-9} Torr. The oxidation process as performed by backfilling the chamber with oxygen to a partial pressure of 400 Torr and controlling the substrate temperature by radiative heating from a wire-wound heater. For oxide growth, the calibrated substrate temperature had an accuracy of ±20 °C over the range used in the experiment (25–600 °C).

The UV light was generated by a Hg lamp (quartz tube) designed to cover a 4-in wafer was placed in close proximity to the substrate. The lamp emitted a 184.9 nm line, which causes the reaction $O_2 + h\nu \rightarrow 2O$ to form oxygen radicals (these also participate in a competing reaction with O_2 to form O_3 molecules), and a 253.7 nm line, which is strongly absorbed by ozone to cause the reaction $O_3 + h\nu \rightarrow O_2 + O$. These processes form a dynamic equilibrium of reactive, oxidizing

---

Manuscript received May 22, 1998; revised November 6, 1998. This work was supported in part by DARPA Contract F49620-96-C-0006.

G. D. Wilk is with the Central Research Laboratories, Texas Instruments, Dallas, TX 75243 USA.

B. Brar is with the Applied Research Laboratories, Raytheon TI Systems, Dallas, TX 75243 USA.

Publisher Item Identifier S 0741-3106(99)02470-2.

0741-3106/99$10.00 © 1999 IEEE

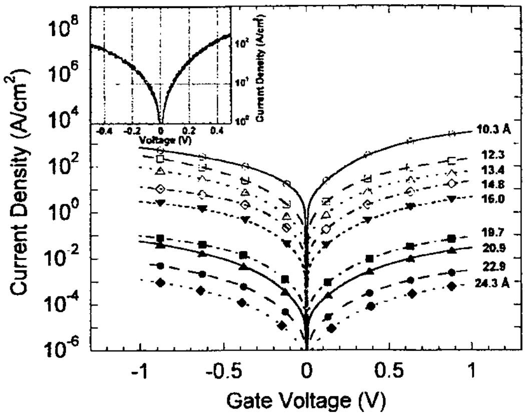


Fig. 1. Current–voltage (I–V) curves for a set of ultrathin oxides formed by UVO3 exposure with different thicknesses. The oxides are well-behaved and show the expected exponential dependence of tunneling current density on thickness. The inset shows I–V curves taken from center to edge across a wafer with an oxide grown at 25 °C. The oxide thickness variation (as determined by I–V) across the region is less than 0.1 Å.

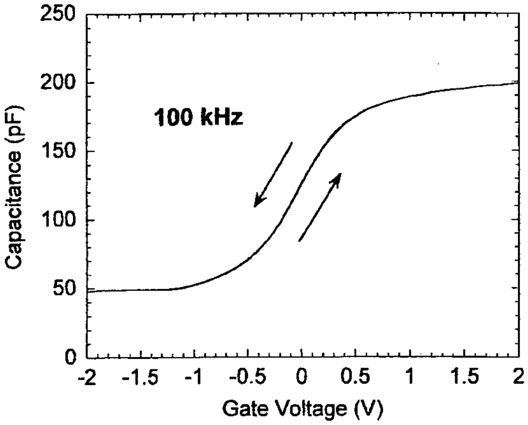


Fig. 2. High-resolution TEM image of an oxide grown by the UVO$_3$ process at 600 °C for 60 min. The oxide is uniform and homogeneous. The thickness is 24 Å, as measured by TEM.

species. Since atomic O is also formed by the thermal dissociation of O_3 molecules into O_2 and O at temperatures above 100 °C, the smaller and more reactive atomic O species most likely are the predominant oxidants [12] in forming SiO_2. X-ray photoelectron spectroscopy measurements were made on the films and showed the oxides to be stoichiometric SiO_2, with the expected peaks at 99.5 eV (Si^0, bulk Si) and at about 103.5 eV (Si^{4+}).

Capacitors were fabricated with electron-beam evaporated Al contacts, and current-voltage (I–V) and capacitance-voltage (C–V) measurements were performed on circular devices with diameters ranging from 15 to 1500 μm. Capacitance-voltage measurements were performed on an HP 4284 LCR meter at frequencies up to 1 MHz. The measured quantities were reproducible, showed no hysteresis or dispersion and scaled linearly with the device area, thereby showing excellent uniformity of the oxide layers.

Fig. 1 shows a set of I–V curves in the direct tunneling regime, taken from ultrathin oxides grown by our UVO$_3$ process with thicknesses ranging from about 10 to 25 Å. The oxide thicknesses were determined by solving a simple transcendental equation relating the direct tunneling current to the oxide thickness, using parameters for ultrathin oxides [9]. The thickness is determined at each bias point measured, and is constant to within ±0.3 Å from 0.01 to 1.0 V. The exponential dependence of tunneling current on oxide thickness is illustrated in Fig. 1, where a decrease of ~2 Å in thickness results in an order of magnitude increase in current density. Spectroscopic ellipsometry (SE) was also performed on each of the oxides. There is good agreement between the oxide thickness values as determined by SE and by I–V measurements for all temperatures and times used.

The inset in Fig. 1 shows a set of I–V curves, measured from center to edge on a 4-in wafer, of a ~10 Å oxide grown at 25 °C (bottom curve in Fig. 1). This plot demonstrates the extremely high oxide uniformity attainable with the UVO$_3$ process, as all of the I–V traces overlay each other very well, representing a thickness variation of less than 0.1 Å (less than 1% standard deviation). The high thickness uniformity is a consequence of the self-limiting nature of the ozone oxidation process.

The thickest oxide grown in this study was used for C–V analysis, as shown in Fig. 2. Higher tunneling currents in the thinner oxides precluded similar C–V analysis of those samples. The C–V curves were measured at 1 MHz in both sweep directions. The oxide thickness is determined to be 24.7 Å, using an analytical C–V method [13]. The plot shows that there is no measurable hysteresis in the capacitance, indicating that it is free of trapped charge. Capacitance–voltage measurements were also made at 10 kHz and showed identical behavior, indicating that there is no dispersion in the oxide film. The flatband voltage is small, $V_{FB} < -40$ mV, which indicates very little oxide charge. This compares favorably to a much larger $V_{FB} \sim -1$ V previously measured on a UVO$_3$-generated oxide, which most likely had a large amount of oxide charge [12].

Fig. 3 is a plot of oxide thickness, as determined by I–V measurements, as a function of substrate temperature and UV ozone exposure time. The dashed lines for times less than 30 min are not measured values, but indicate that some finite time is required before the nearly self-limiting thickness is reached. Although the initial ~10 Å grows rapidly at the onset of the oxidation, subsequent oxidation between 30 and 60 min occurs slowly. The slight temperature dependence (for $T < 600$ °C) of the oxidation rate is most likely attributable to the temperature dependence of the diffusivity of atomic O through the oxide. This weak temperature dependence allows highly accurate growth of oxide films in the ultrathin regime. For example, to grow a ~20 Å oxide, one could choose conditions of 550 °C and 400 Torr O_2 for 30 min. Since the oxidation rate under these conditions is about 0.05 Å/min, the oxidation time of 30 min allows for large time variations and thus more flexibility in the oxidation process. The oxide grown at 600 °C for 60 min was determined from I–V measurements to be 24.3 Å thick, as opposed to 24.7 Å as obtained by C–V analysis after Maserjian *et al.* [13].

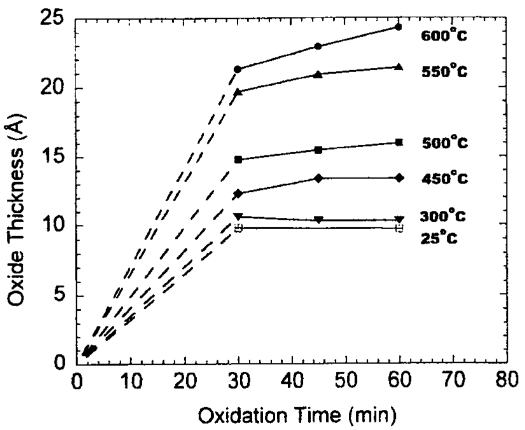


Fig. 3. Capacitance–voltage ($C-V$) curves of both positive and negative sweeps at 1 MHz on the oxide shown in Fig. 2. There is no measurable hysteresis in this oxide. The oxide thickness as measured by $C-V$ analysis is 24.7 Å.

The slopes of the curves for oxidation times between 30 and 60 min in Fig. 3 increase with increasing temperature, indicating a contribution from thermal oxidation to the overall oxide thickness. To determine the contribution of thermal oxidation (i.e. by O_2 rather than atomic O) to the total oxide thickness over these long anneal times, several oxidations were carried out under the same conditions as shown in Fig. 3, except that the Hg lamp was turned off. The pure thermal oxidation rates at temperatures of 550 °C and 600 °C were significantly lower than the slopes of the respective curves in Fig. 3, in agreement with previous results [12]. This indicates that the combination of thermal and ozone oxidation leads to a greater oxidation rate than just by thermal oxidation. A more detailed discussion of these effects, along with the oxidation kinetics of the UVO_3 process, will be discussed in a later publication [14].

The UVO_3 oxidation process allows uniform exposure of O_3, O and O_2 molecules to the surface of the Si wafer. Since the process is nearly self-limiting in thickness, the interfacial growth front does not continually move into the substrate, and this allows uniform formation of the ultrathin oxide at temperatures below 600 °C. A slight temperature dependence of the oxidation rate above 450 °C does cause an increase in thickness for longer exposure times. This time dependence is weak, however, and can be easily accounted for by choosing oxidation conditions for a desired oxide thickness. The ability to obtain precise oxide thicknesses with very high uniformity in the ultrathin regime, and at a lower thermal budget than normally used for thermal oxidation make the UVO_3 process potentially very useful for scaled CMOS devices.

ACKNOWLEDGMENT

The authors are grateful to M. Howell and W. Dunn for excellent technical support, and to L. Magel for the XPS measurements.

REFERENCES

[1] H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S.-i. Nakamura, M. Saito, and H. Iwai, "1.5-nm direct-tunneling gate oxide Si MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, p. 1233, 1996.

[2] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S.-H. Lo, G. A. Sai-Halasz, R. G. Viswanathan, H.-J. C. Wann, S. J. Wind, and H.-S. Wong, "CMOS scaling into the nanometer regime," *Proc. IEEE*, vol. 85, p. 486, 1997.

[3] C. Wann, K. Noda, T. Tanaka, M. Yoshida, and C. Hu, "A comparative study of advanced MOSFET concepts," *IEEE Trans. Electron Devices* vol. 43, p. 1742, 1996.

[4] G. D. Wilk, "Surface preparation, growth and characterization of ultrathin gate oxides for scaled CMOS applications," *SPIE Proc.*, vol. 3212, p. 42, 1997.

[5] L. Manchanda, "Gate dielectrics for high performance ULSI: Practical and fundamental limits," in *Semiconductor Characterization: Present Status and Future Needs*, W. M. Bullis, D. G. Seiler, and A. C. Diebold, Eds. New York: AIP, 1996, p. 123.

[6] G. D. Wilk, Y. Wei, H. Edwards, and R. M. Wallace, "*In situ* flux cleaning technique for producing atomically flat Si(100) surfaces at low temperature," *Appl. Phys. Lett.*, vol. 70, p. 2288, 1997.

[7] G. Timp, *et al.*, "Low leakage, ultrathin gate oxides for extremely high performance sub-100-nm nMOSFET's," in *IEDM Tech. Dig.*, 1997, p. 930.

[8] B. Brar, G. D. Wilk, and A. C. Seabaugh, "Direct extraction of the electron tunneling effective mass in ultrathin SiO_2," *Appl. Phys. Lett.*, vol. 69, p. 2728, 1996.

[9] T. Nakanishi, S. Ohkubo, and Y. Tamura, "Oxidation in ozone," *Fujitsu Sci. Tech. J.*, vol. 32, p. 128, 1996.

[10] V. Nayar, P. Patel, and I.W. Boyd, "Atmospheric pressure, low temperature (<500 °C) UV/ozone oxidation of silicon," *Electron. Lett.*, vol. 26, p. 205, 1990.

[11] T. Nakanishi, Y. Sato, M. Okuno, and K. Takasaki, "Improvement in MOS reliability by oxidation in ozone," in *VLSI Tech. Dig.*, 1994, p. 45.

[12] A. Kazor and I. W. Boyd, "Ozone-induced rapid low temperature oxidation of silicon," *Appl. Phys. Lett.*, vol. 63, p. 2517, 1993.

[13] J. Maserjian, G. Petersson, and C. Svensson, "Saturation capacitance of thin oxide MOS structures and the effective surface density of states of silicon," *Solid-State Electron.*, vol. 17, p. 335, 1974.

[14] To be published.

SURFACE PREPARATION, GROWTH AND CHARACTERIZATION OF ULTRATHIN OXIDES

Glen Wilk, Berinder Brar*, Yi Wei, Robert M. Wallace and Alan Seabaugh*
Materials Science Laboratory
and
**System Components Laboratory*
Corporate R&D, Texas Instruments, Dallas, TX 75265

ABSTRACT

We report on recent experimental results on several aspects of surface preparation and ultrathin oxide growth on Si(100) from an atomic level perspective. A new, *in situ* surface cleaning method which yields very smooth, well-ordered surfaces is compared to a standard HF-based etching technique, with emphasis on the surface morphology as revealed by atomic force microscopy and scanning tunneling microscopy. Using this technique, ultrathin oxides grown on these surfaces show improved uniformity and reproducibility. Structural and electrical measurements on single layer ultrathin oxides will be reported, and double layer oxide structures are also discussed. The problem of accurate oxide thickness measurements in the ultrathin regime will be addressed along with recent data comparing several different measurement techniques.

1. Introduction

The continued reduction in gate oxide thickness as demanded by CMOS scaling has fueled many studies on the properties of thin oxides. According to the SIA Roadmap[1], by the year 2004, 0.13 µm technology will require a 3.4 nm thick oxide layer to maintain high current drive in MOSFET devices. At this thickness, significant direct tunneling occurs between the gate and the substrate, even at low applied biases. In addition, for ultrathin oxides on this scale, the contribution of the SiO_2/Si interface to oxide reliability and quality is greatly magnified. It is clear that features on the atomic scale will profoundly influence the properties of ultrathin oxides. These concerns have driven researchers to study the detailed structure of the SiO_2/Si interface as well as consider altogether different, alternate dielectrics to replace SiO_2.

In order to better understand the nature of the interface between Si and SiO_2, both physical and electrical characterization are necessary. Using MOS structures, electrical measurements can yield a wealth of information about the electronic nature of the interface while imaging techniques can be used to elicit the physical nature of the interface. The combination of these analyses will greatly improve our understanding of these interfaces, and with better control, perhaps will lead to acceptable ultrathin gate oxides.

The decrease of oxide thicknesses into the sub-3.0 nm regime also presents a serious problem of being able to accurately and reproducibly measure the oxide thickness. Spectroscopic ellipsometry is currently the standard technique, but it is unreliable for films below 3.0 nm. Considering that only very small variations in oxide thickness across the wafer will be tolerable (due in part to the exponential dependence of tunneling current on oxide thickness), a better understanding of the interface on the atomic scale is needed. A comprehensive study of the available techniques for measuring these thin films is required to establish a standard for measuring these thicknesses.

Some structures such as resonant tunneling diodes (RTDs), which are used in nanoelectronics applications, actually prefer a barrier layer which is as thin as possible. Since these devices rely on electron transmission for high current output, thinner layers are required to obtain good performance. We are investigating oxides less than 1.0 nm thick for potential use in Si-based RTDs through highly controlled oxidation and Si growth conditions.

2. Si Flux Cleaning

The ultrahigh vacuum (UHV) system used for all sample fabrication on 4-inch wafers in this paper is shown in Fig. 1. The molecular beam epitaxy (MBE) chamber is used for deposition of Si, Ge, Sb, B and CaF_2 from electron beam evaporators and K-cells, and is equipped with RHEED for monitoring real-time evolution of the surface. A substrate heater and rotating stage allow for uniform, variable-temperature depositions. The wafers can be transferred via a trolley system from one chamber to another while remaining under UHV. Some oxidations are done in the sputter chambers, one of which contains a substrate heater, so that the relatively high pressures used for oxidation do not interfere with the operations of the MBE chamber.

The two sputter targets can be used for deposition of many materials which are not suitable for evaporation in the MBE chamber. A separate apparatus in one of the loadlocks allows ultrathin oxidation by UV ozone exposure (from a Hg lamp) at room temperature. Following oxidation, the wafers are transferred to the metal deposition chamber where Al electrode top layers can be deposited without breaking vacuum. This versatile UHV MBE deposition system allows for numerous *in situ* studies of many different materials structures.

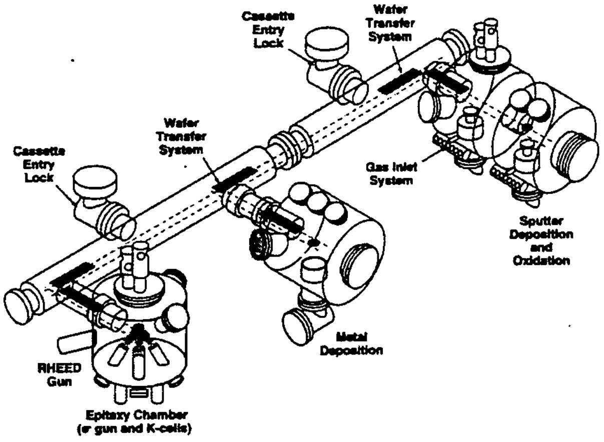

Cassette Entry Lock
Wafer Transfer System
Wafer Transfer System
Cassette Entry Lock
Gas Inlet System
Sputter Deposition and Oxidation
Metal Deposition
RHEED Gun
Epitaxy Chamber (e- gun and K-cells)

Figure 1. UHV MBE system for *in situ* cleaning and growth of ultrathin oxides and alternate dielectrics.

Si surfaces are typically prepared by using HF to etch the surface oxide, leaving a smooth, H-terminated Si surface. While this method does not require high temperatures, it yields a featureless surface. An alternative technique that we have developed in MSL does not require *ex situ* chemical cleaning and yields an atomically smooth surface with a well-defined step structure. The method involves use of a Si flux from an electron beam evaporator, which impinges on the sample surface (which contains an oxide) as the sample is heated up to ~ 800°C. Two mechanisms for oxide removal are illustrated in Fig. 2. Both involve a reaction between Si monomers and SiO_2 molecules which forms volatile SiO molecules that desorb from the sample[2]. As shown in Fig. 2(a), a substrate with an oxide is heated in vacuum (a standard procedure) to 700 - 900°C for several minutes.

The above etching reaction occurs at the SiO_2/Si interface and holes form in the Si substrate. In order to obtain a high-quality, smooth and well-ordered surface, the holes must be prevented from forming. To achieve this with the Si flux technique, it is critical that a Si flux be initiated at the *onset* of the substrate heating.

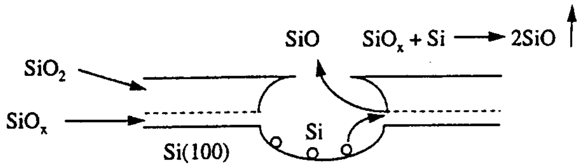

SiO
$SiO_x + Si \rightarrow 2SiO\uparrow$
SiO_2
SiO_x
Si
Si(100)

Figure 2(a). Reaction for oxide removal at the SiO_2/Si(100) interface caused by heating the substrate above 700°C in vacuum. Large holes form in the substrate.

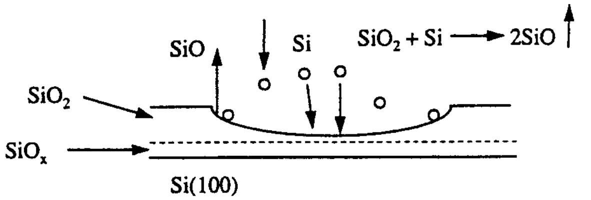

SiO
Si
$SiO_2 + Si \rightarrow 2SiO\uparrow$
SiO_2
SiO_x
Si(100)

Figure 2(b). Reaction at the vacuum/SiO_2 interface with a Si flux impinging on the surface above 700°C. The oxide is removed but no holes are formed.

In this way, the oxide is removed by Si monomers from the molecular beam source and no holes are formed in the substrate, as seen in Fig. 2(b). This process is continued for times ranging from 10 - 30 minutes.

The resulting surfaces, cleaned under these conditions, are extremely smooth and well-ordered. Figure 3(a) shows an atomic force microscopy (AFM) image of a Si(100) surface which has been cleaned by an HF vapor method[3]. The measured rms roughness is 0.1 nm, which is quite smooth, but there are no distinctive

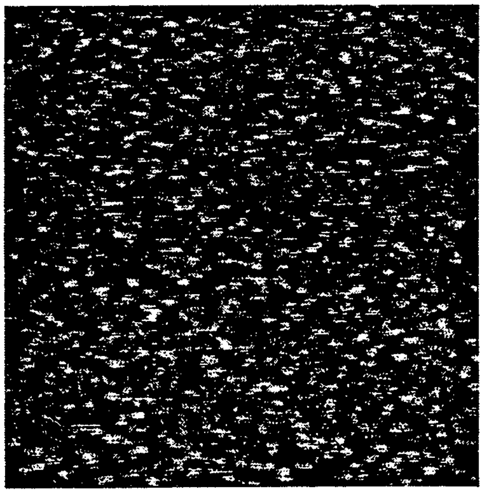


Figure 3(a). 1μm x 1μm AFM image of an H-terminated Si(100) surface. The surface is smooth (rms ~ 0.1 nm) but no steps are visible.

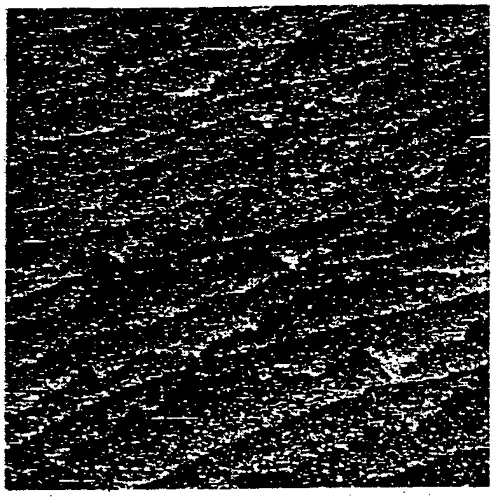


Figure 3(b). 500 nm x 500 nm STM image of a flux cleaned Si(100) surface. The surface is very smooth with well-defined steps. Terrace width is ~ 25 nm.

features visible. The scanning tunneling microscopy (STM) image in Fig. 3(b) shows that using the Si flux cleaning technique on the same substrate, an atomically flat surface is obtained with regular, single atomic height steps. Both A- and B-type steps are clearly visible, as a further indication of the high-quality surface. The measured rms roughness is 0.05 nm with 0.15 nm step heights. These surfaces allow subsequent overgrowth of high-quality films. Smooth interfaces are important and perhaps necessary for obtaining improved reliability in ultrathin gate oxides.

3. Ultrathin Oxides

Electrical characterization of Al/SiO$_2$/n$^+$ Si MOS structures with thermal oxide films ranging from 1.6 - 3.5 nm thick (as determined by C-V measurements) is shown in Fig. 4. The device areas are ~ 1 x 10^{-5} cm^2. Figure 4(a) shows the I-V measurements for direct tunneling in the oxide films, and it can be seen that the curves are all similar in shape and slope. The current increases exponentially with decreasing oxide thickness, as expected for direct tunneling. All of these features indicate very uniform, reproducible oxide films. Figure 4(b) shows the C-V curves (measured at 100 kHz) for the samples in 4(a), and it can be seen that they correspond well with the I-V curves for the same device. The C-V curves also behave very nicely, as the capacitance in the accumulation regime saturates quickly, with parallel slopes among the various curves. The uniformity and reproducibility of these measurements demonstrates the high degree of uniformity and reproducibility of the oxides. Note that the C-V curve for the 1.6 nm oxide is unstable even at low bias. This occurs because the tunneling current is comparable to the displacement current used in making the measurement. This phenomenon precludes high frequency C-V measurements of oxides much below 2.0 nm thick.

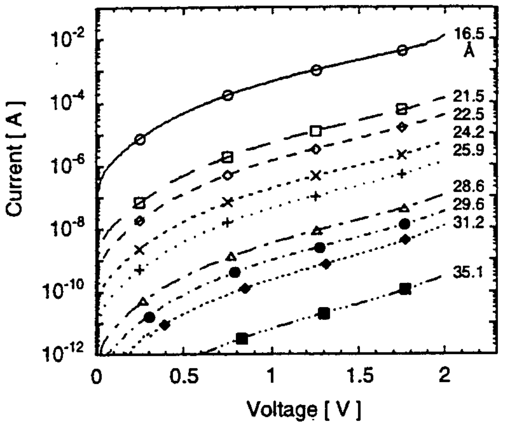


Figure 4(a). I-V curves show very good uniformity and well-behaved exponential dependence of current on oxide thickness.

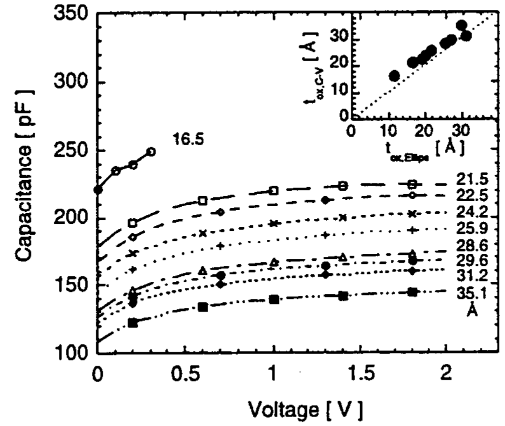


Figure 4(b). C-V curves of the same oxides as in 4(a). The capacitance behavior is uniform and the corresponding oxides show very similar properties.

Since these oxides are very uniform and reproducible, however, an Arrhenius-type graph of current vs. oxide thickness allows extrapolation of the oxide thickness far below 2.0 nm. Oxides grown by UV ozone oxidation, which are self-limiting at ~ 0.9 nm, have been analyzed in this way but are not shown here. In addition, by using the same graphing analysis we have established a method for extracting a value for the electron effective mass in the oxide which is ~ 5x more accurate than previous measurements[4-6]. The value we obtained is 0.30 ± 0.02m$_e$. Thus this high uniformity has allowed further evaluation of fundamental properties of the oxides. Measurements on these oxides will provide better insight into the potential for using ultrathin gate oxides with direct tunneling currents.

Another application of ultrathin oxides is in nanoelectronics for fabricating RTDs. In these devices, thinner barriers are preferable for obtaining large current flows under resonance conditions. It is desirable to have two barriers separated by a quantum well, in which quantum levels are created depending on the width of the well. Figure 5 shows a cross-sectional high-resolution TEM picture of a structure which consists of two 0.9 nm SiO_2 layers with a 2.0 nm amorphous Si layer between them. Al is used for the electrode and the substrate is Si(100). Electrical characterization of these structures is currently underway.

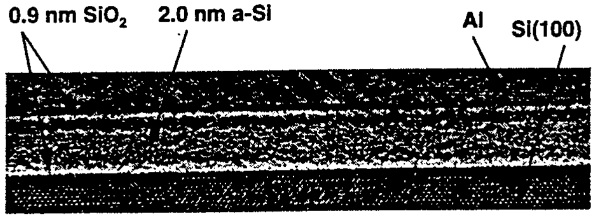


Figure 5. High-resolution cross-sectional TEM of a double barrier SiO_2/a-Si/SiO_2 structure. The barrier layers are flat and uniform.

The image shows the high uniformity obtained by UV ozone oxidation, as opposed to thermal oxidation, and the interfaces are very uniform and flat. These interfaces are necessary for obtaining electron coherence and consequently achieving resonant tunneling. Electrical characterization on these structures is underway.

4. Measurement of Ultrathin Oxide Thickness

As oxide thicknesses decrease to the sub-3.0 nm range, it becomes increasingly important to be able to accurately measure the oxide thickness. The most often used techniques are ellipsometry, C-V and XPS, but each type of analysis has disadvantages and uncertainties which need to be better understood and quantified with regard to oxide thickness. Figure 6 lists measurements made on a set of ultrathin oxides using facilities in MSL. It can be seen that there are significant discrepancies among these three techniques. This is somewhat expected, however, because each technique measures a different property of the oxides. An important task is to correlate and quantify all of these differences and be able to reliably interpret the analysis of one technique in terms of another.

Ellips (A)	C-V/I-V (A)	XPS (A)
32.8	30.9	27.0
29.7	35.1	26.1
21.5	25.9	24.5
15.7	15.5	10.6
11.5	16.5	11.2
7.7	10.3	7.3

Figure 6. Table of oxide thicknesses (in Angstroms) as measured by ellipsometry, C-V, I-V and XPS.

In the case of ellipsometry, very thin oxide layers make the measurements unreliable because of contributions from interface roughness, contamination and lack of knowledge of optical constants in this thickness regime. C-V measurements electrically probe the region over which charge is held, but for semiconductor substrates, a depletion region extends into the substrate which supports a significant amount of charge. This tends to increase the "electrical" thickness of the layer. As mentioned before, C-V measurements are not possible for oxide thicknesses much below 2.0 nm because of quantum tunneling effects. These oxide thicknesses (below 2.0 nm) were extrapolated based on I-V measurements, as mentioned previously. XPS yields a very sensitive measurement of the amount of oxide, but this method assumes that the film morphology is very uniform and it cannot easily account for the contribution of suboxides (SiO_x, $0<x<2$) even though they comprise a non-negligible portion of the total layer in this regime. TEM is another possibility but this approach suffers from being destructive, time-consuming and because it is a very local measurement. Further work in comparing and correlating these techniques will provide a better understanding for making accurate thickness measurements of ultrathin oxides.

5. References

[1] *National Technology Roadmap for Semiconductors*, SIA, (1994).
[2] G.W. Rubloff, *J. Vac. Sci. Technol. A* **8**, 1857 (1990).
[3] R. Wise, D. Frystak, J. Barnett, P. Grothe, B. Fowler and E. Forest, *TI TAR*, December, 1995.
[4] J. Maserjian, *J. Vac. Sci. Technol. A* **11**, 996 (1974).
[5] M. Hiroshima, T. Yasaka, S. Miyazaki and M. Hirose, *Jap. J. Appl. Phys. Pt. 1* **33**, 395 (1994).
[6] M. Depas, R.L. van Meirhaeghe, W.H. Laflere and F. Cardon, *Solid State Electronics* **37**, 433 (1994).

In situ Si flux cleaning technique for producing atomically flat Si(100) surfaces at low temperature

G. D. Wilk,[a] Yi Wei, Hal Edwards, and R. M. Wallace
Central Research Laboratories, Texas Instruments, Dallas, Texas 75265

(Received 4 November 1996; accepted for publication 28 February 1997)

We have developed a method for removing oxides and producing atomically flat Si(100) surfaces with single atomic height steps using a Si flux cleaning technique. By introducing a Si flux in the range 1.0–1.5 Å/s at the onset of an SiO_2 thermal desorption step as low as 780 °C, scanning tunneling microscopy (STM) and atomic force microscopy images reveal smooth surfaces with atomically flat terraces with an rms roughness of 0.5 Å and single-step heights of 1.4 Å. STM reveals that A- and B-type steps are present across the entire area of the scanned surface. Desorption of the surface oxide layer with Si fluxes below this range results in rougher surfaces with pits ~50 Å deep and 1000 Å across. For Si fluxes above this range, no pits are seen but atomic steps are not visible on the surface. © *1997 American Institute of Physics.* [S0003-6951(97)04017-5]

Device scaling in CMOS is requiring ever thinner gate oxides while still demanding high reliability of these oxides. The gate oxide thinning trend is now placing more focus on the interface between the SiO_2 layer and the Si substrate as a means of identifying the sources of premature breakdown and perhaps improving reliability.[1–3] The interface roughness depends strongly on the cleaning procedure used in preparing the Si surfaces, and standard techniques[4,5] of producing a surface oxide cap followed by an *in situ* anneal step at 800–900 °C can actually produce pits on the surface which are 50 Å deep and up to 1000 Å across. For thin-film growth, this large roughness can significantly affect the properties of the film.

It has been suggested for thick oxides[6–8] (~200 Å) and for thin oxides[9–12] ($\leqslant$30 Å) that the pits form by SiO etching of the substrate, where suboxides ($SiO_x, 0<x<2$) near the interface desorb as volatile SiO molecules, followed by reduction of the oxide layer in the reaction $SiO_2+Si \rightarrow 2SiO$ (g). Mobile Si monomers are suggested[10] to form in the substrate voids left by the suboxides near the interface which then reduce SiO_2 molecules to the volatile SiO form, which desorbs from the substrate. Engel *et al.*[9] has also argued through energetics considerations that the formation of mobile monomers is the rate-limiting step in the reaction. Recent investigations on the void growth process, however, suggest that mobile Si monomer formation and SiO formation are kinetically competing processes on the Si surface.[13] While this method removes the oxide, it does so at the expense of forming pits in the Si substrate. It is very desirable, however, to remove the oxide without forming pits in the substrate and also without using chemical etching. Removing the chemical etching steps could reduce and simplify the number of surface preparation steps used in standard processing procedures. Further, while HF etching removes the oxides and yields a smooth, H-terminated surface (rms roughness ~1 Å), there is no well-ordered structure to the surface. An atomically flat Si surface with well-ordered steps can be obtained by flashing the sample at 1200 °C but this process is not practical for large wafers and thermal budget concerns.

A typical surface preparation method starts with the formation of a chemical oxide as a protective cap on the Si surface. Once the sample is placed in vacuum, resistive or radiative heating is used to anneal the sample at 800–900 °C for several minutes to desorb the oxide. Various low Si deposition rates (~0.1–0.6 Å/s) have also been used[14–16] for several minutes in an attempt to improve the surface quality for subsequent thin-film growth. While the emergence of reflection high-energy electron diffraction (RHEED) streaks and spot patterns was used as an indication of oxide removal in some of these studies, it was assumed that these processes yield a flat, uniform surface. Under many oxide desorption conditions, however, a rough surface can result which may have a significantly detrimental effect on ultrathin film properties, particularly where a uniform interface is required. To our knowledge no one has investigated the Si surfaces with atomic resolution probes following these treatments. An *in situ* Si surface cleaning technique which uses relatively low temperatures and yields an atomically flat, stepped surface is very desirable. In this letter, we describe a Si flux cleaning technique which can be used to obtain atomically flat, single-height steps on Si(100) at temperatures between 700 and 800 °C. The surface cleaning conditions have been varied as a function of Si deposition rate and substrate temperature.

The substrates were nominally flat 4-in Si(100), *n*-type wafers with ρ=0.01–0.02 Ω cm. Some of the as-received wafers were placed directly into a molecular beam epitaxy (MBE) chamber without any *ex situ* chemical cleaning, while others received an initial degreasing with acetone and methanol. The substrates where heated radiatively by a strip heater near the backside of the wafer. A thermocouple placed near the substrate was used for temperature measurements. Since the substrate holder rotates during the surface preparation procedure, direct contact between the thermocouple and substrate was not possible. The thermocouple was calibrated to an optical pyrometer over the temperature range used in this experiment. While the temperature reproducibility was ±5 °C, the absolute temperature is accurate to ±25 °C. After a long degas at 400 °C, the wafers were heated up to 650 °C, at which point the wafer surface was exposed to the Si flux. The anneals were then done at various temperatures and times while the Si flux impinged on the substrate. The base pressure in the MBE chamber was 5×10^{-11} Torr, and was $\sim8\times10^{-10}$ Torr during Si deposition. Following the Si flux

---

[a] Electronic mail: wilk@crlmsm.csc.ti.com

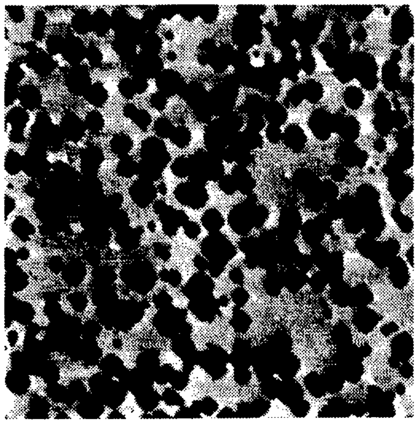


FIG. 1. 10×10 μm^2 AFM image of Si(100) with a native oxide which has been partially desorbed by annealing at 780 °C for 30 min. A high density of voids (1000–5000 Å wide, 40 Å deep) is seen on the surface.

cleaning step, the wafers were removed from vacuum, cleaved and transferred in air as rapidly as possible (no longer than 30 min) for STM (in UHV after transfer) and atomic force microscopy (AFM) (in ambient, performed in ''tapping'' mode) analyses.

Figure 1 shows a 10×10 μm^2 ambient AFM image of a Si(100) surface which has undergone a partial thermal desorption of the native oxide. The sample was heated at 780 °C for 30 min under UHV conditions. A very high density of holes is clearly seen across the surface. The holes range from ~500 to 5000 Å wide and are ~40 Å deep. The initial native oxide is still present in the regions between the holes. With no Si flux provided during the anneal step, it is seen that a rough, nonuniform surface results.

Figure 2(a) shows a 10×10 μm^2 AFM image of a Si(100) surface (initially containing a native oxide) which has been annealed at 780 °C for 30 min, with a Si flux (~0.3 Å/s of Si) being introduced for the *final* 10 min of the anneal only. While etching of the substrate still occurs, the void density and size have been reduced by the addition of a Si flux during the anneal. Figure 2(b) shows a 5000 ×5000 Å^2 UHV STM image of the sample in Fig. 2(a). Compared to the surface in Fig. 1, the surface is much smoother between the voids, and the voids themselves are much better defined: there is clearly a four-fold symmetry to the voids, reflecting the symmetry of the Si(100) substrate. The addition of a Si flux has resulted in a clean surface and the formation of a well-defined step structure corresponding to the original miscut angle of the surface. This indicates that although the addition of a Si flux for the final 10 min helps to reduce the void size and density, it does *not* prevent void formation. The holes also appear to act as step pinning centers, as the image shows several steps with very small terrace widths near the hole edge. This behavior may contribute to the depth of the holes in addition to substrate etching by SiO formation.

Figure 3 shows a 1×1 μm^2 STM image of a Si(100) surface (initially containing a native oxide) which has been annealed at 780 °C for 10 min, with the Si flux (~1 Å/s) being introduced at the *onset* of the thermal anneal (at 650 °C, followed by a temperature ramp up to 780 °C). No holes are visible (larger-area scans show no holes across 100 μm^2 in AFM). The measured rms roughness is 0.5 Å for this image. Alternating A-type (smooth) and B-type (rough) steps can be seen in the image which indicates a smooth, high quality surface with 1.4 Å single atomic height steps. This image reveals an extremely well-defined step structure on the surface. Atomically flat surfaces such as this one were observed for all samples which were Si flux-cleaned with the Si deposition rate between 1.0 and 1.5 Å/s.

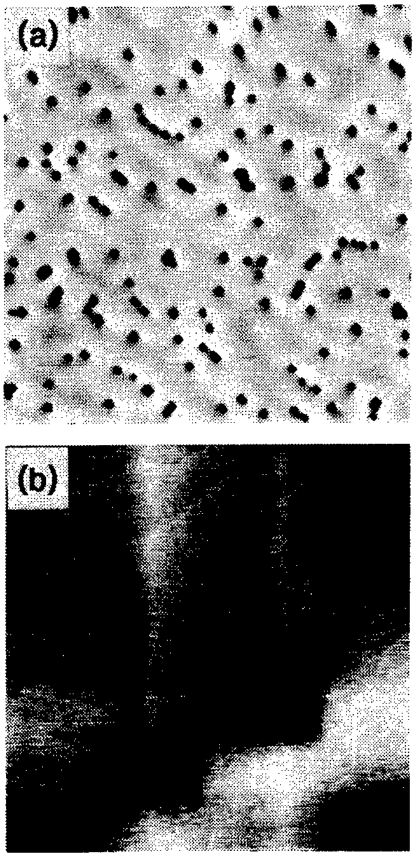


FIG. 2. (a) 10×10 μm^2 AFM image of Si(100) after complete desorption of the native oxide. The sample was annealed at 780 °C for 30 min with a Si deposition rate=0.3 Å/s being introduced for the *final* 10 min of the anneal. (b) 0.5×0.5 μm^2 STM image of the sample in Fig. 2(a). The holes reflect the four-fold symmetry of the substrate and act as step pinning centers.

Most Si cleaning methods which use Si deposition to improve the surface quality introduce a Si flux at some point after the heating has been initiated, and often it is done after the anneal step has been completed. This work demonstrates, however, that it is critical that the Si flux be started at the *onset* of the heating step, in our case 650 °C, so that reduction of the oxide at the top surface prevents oxide reduction at the interface. It is also essential that Si is deposited at an appropriate rate to compete with the interfacial reduction reaction. For Si deposition conditions outside of the range of 0.2–2 Å/s, the surface becomes significantly rougher. For Si deposition rates below ~0.2 Å/s, Si is not supplied quickly enough to reduce SiO_2 from the top surface, and voids form from reaction at the bottom SiO_2 interface. For deposition rates higher than ~2 Å/s, Si deposits too quickly for the decomposition to occur, thus excess Si blankets the surface and creates less well-defined steps. It is also possible that not all of the SiO_2 is desorbed in this case, which would account

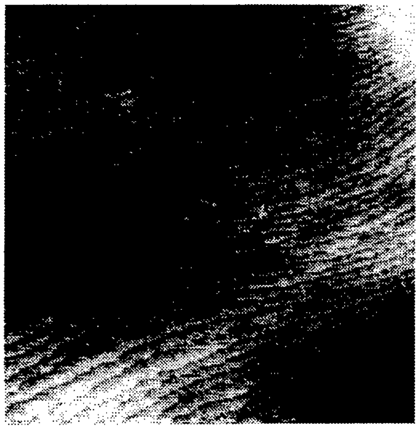


FIG. 3. 1×1 μm^2 STM image of Si(100) after desorption of the native oxide at 780 °C for 10 min with a Si flux of ~1 Å/s, introduced at the *onset* of the thermal desorption step. The measured rms roughness is 0.5 Å and both A-type and B-type single atomic height steps can be seen.

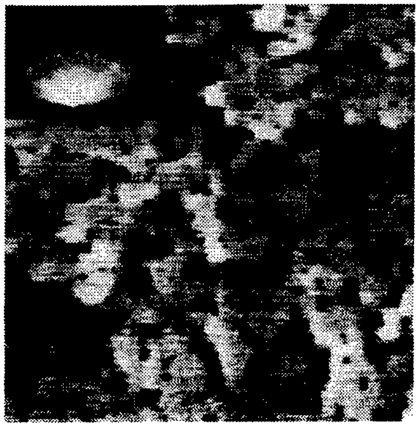


FIG. 4. 1×1 μm^2 STM image of a Si sample initially containing a native oxide after an anneal at 800 °C for 10 min with a Si flux of 6 Å/min. A highly nonuniform surface results with a high density of large holes. The inset shows the corresponding RHEED pattern taken from this surface.

for the less well-defined surface structure. The optimum Si deposition rates were found to be 1.0–1.5 Å/s, where STM and AFM both observed atomically flat terraces and single-height A- and B-type steps (as seen in Fig. 3). Within this range, it appears that the Si monomer arrival rate is balanced with the SiO_2 reduction rate and a high-quality surface can be obtained.

These results indicate that at low deposition rates a rough, very nonuniform surface results. We reproduced previously reported cleaning conditions by preparing a sample with an anneal step at 800 °C for 10 min and a Si flux of ~6 Å/min, which closely approximates the rate of ~8 Å/min at 800 °C as suggested by others.[15] Figure 4 shows a 1 ×1 μm^2 STM image of the surface under these conditions. It can be seen that the surface is marked by a very high density of pits which are ~50 Å deep. The inset of Fig. 4 shows the corresponding RHEED pattern from this surface, which contains streaks normally associated with a smooth, reconstructed crystalline surface. Although RHEED is often used as an indicator for smooth, uniform and crystalline surfaces, defects of this type will yield a deceptive RHEED pattern which may be mistaken for a high-quality surface. The presence of pits is expected only to decrease the intensity of and slightly broaden the RHEED spots.[17] It is clear that this type of RHEED pattern may be incorrectly interpreted without STM examination of the surface. These surface nonuniformities can have a profound effect on any subsequent film growth.

At substrate temperatures up to 900 °C, the surface maintains sharp single-height steps and atomically flat terraces. This behavior is not surprising, as an increased Si adatom diffusion length maintains high surface quality. At much higher temperatures, however, it is expected that surface roughening will occur during deposition. High-quality surfaces have also been observed down to temperatures of 700 °C, although the step edges are not as well defined and the terraces show more roughness. Within large windows of deposition rate and substrate temperature, very high-quality Si surfaces can be obtained using this technique.

This method appears to work well because the oxide reduction reaction at the surface has one less reaction step than that at the interface. Since the Si atoms arrive uniformly as monomers in the flux, the activation barrier for Si monomer formation is removed. Recent calculations suggest[18] that the activation energy for Si monomer formation from the 2×1 reconstructed surface is comparable to that for reduction of SiO_2 to SiO. If this is the case, the reaction at the oxide surface may occur more rapidly than the interface reaction. For very thin oxides ($\leqslant$30 Å), the SiO_2 layer can be completely removed before voids of significant size form at the interface. For thicker oxides (>30 Å), significant reduction of the oxide may be unavoidable by this technique, but the Si flux cleaning method should work well for oxides formed in air and by standard chemical cleaning procedures.

The authors thank Marshall Howell and John Powell for their excellent technical assistance and J. P. A. van der Wagt for very useful discussions.

[1] K. F. Schuegraf and C. Hu, Semicond. Sci. Technol. **9**, 989 (1994).
[2] C. J. Sofield and A. M. Stoneham, Semicond. Sci. Technol. , 215 (1995).
[3] H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S.-I. Nakamura, M. Saito, and H. Iwai, IEEE Trans. Electron Devices **43**, 1233 (1996).
[4] Ishizaka, N. Nakagawa, and Y. Shiraki, *Proceedings of the 2nd International Symposium on Molecular Beam Epitaxy* (Japanese Society of Physics, Tokyo, Japan, 1982), p. 182.
[5] W. Kern, *Handbook of Semiconductor Wafer Cleaning Technology* (Noyes, Park Ridge, NJ, 1993).
[6] R. Tromp, G. W. Rubloff, P. Balk, F. K. LeGoues, and J. van Loenen, Phys. Rev. Lett. **55**, 2332 (1985).
[7] M. Liehr, J. E. Lewis, and G. W. Rubloff, J. Vac. Sci. Technol. A **5**, 1559 (1987).
[8] G. W. Rubloff, J. Vac. Sci. Technol. A **8**, 1857 (1990).
[9] Y.-K. Sun, D. J. Bonser, and T. Engel, Phys. Rev. B **43**, 14309 (1991).
[10] Y.-K. Sun, D. J. Bonser, and T. Engel, J. Vac. Sci. Technol. A **10**, 2314 (1992).
[11] K. E. Johnson and T. Engel, Phys. Rev. Lett. **69**, 339 (1992).
[12] K. E. Johnson, P. K. Wu, M. Sander, and T. Engel, Surf. Sci. **290**, 213 (1993).
[13] Y. Wei, R. M. Wallace, and A. C. Seabaugh, Appl. Phys. Lett. **69**, 1270 (1996).
[14] M. Tabe, K. Arai, and H. Nakamura, Jpn. J. Appl. Phys. **20**, 703 (1981).
[15] M. Tabe, Jpn. J. Appl. Phys. 1 **21**, 534 (1982).
[16] K. Kugimiya, Y. Hirofuji, and N. Matsuo, Jpn. J. Appl. Phys. 1 **24**, 564 (1985).
[17] J. E. Mahan, K. M. Geib, G. Y. Robinson, and R. G. Long, J. Vac. Sci. Technol. A **8**, 3692 (1990).
[18] S. Tang (private communication).

Direct extraction of the electron tunneling effective mass in ultrathin SiO_2

B. Brar,[a] G. D. Wilk, and A. C. Seabaugh
Corporate Research Laboratories, Texas Instruments, Inc., Dallas, Texas 75243

(Received 14 May 1996; accepted for publication 22 August 1996)

Electron transport in ultrathin (t_{ox}<40 Å) Al/SiO_2/$n-$Si structures is dominated by direct tunneling of electrons across the SiO_2 barrier. By analyzing the tunneling currents as a function of the SiO_2 layer thickness for a comprehensive set of otherwise identical samples, we are able to extract an effective mass for the tunneling electron in the SiO_2 layer. Oxide films 16–35 Å thick were thermally grown *in situ* in a dry oxygen ambient. The oxide thicknesses were determined by capacitance–voltage measurements and by spectroscopic ellipsometry. The tunneling effective mass was extracted from the thickness dependence of the direct tunneling current between an applied voltage of 0 and 2 V, a bias range that has not been previously explored. Employing both a parabolic and a nonparabolic assumption of the $E-\kappa$ relationship in the oxide forbidden gap, we found the SiO_2 electron mass to be $m_P^*=0.30\pm0.02m_e$, and $m_{NP}^*=0.41\pm0.01m_e$, respectively, independent of bias. Because this method is based on a large sample set, the uncertainty in the mass determination is significantly reduced over prior current-voltage fitting methods. © *1996 American Institute of Physics.* [S0003-6951(96)04644-X]

The properties of silicon dioxide (SiO_2) have been studied since ancient times, but never more intensely than in the past three decades. In metal/oxide/semiconductor (MOS) transistors, the ability to decrease the oxide thickness has allowed device sizes to be reduced while still maintaining high current drive. Oxide thicknesses in the range 50–100 Å are commonplace today in manufacturing, and research into sub-50 Å oxides has intensified. Below 50 Å, the quantum mechanical tunneling current becomes a limiting leakage component in MOS technology, and an enabling element in resonant tunneling and light-emitting silicon devices.[1,2] Toward this end it is important to have reliable measurements of the basic tunneling properties of ultrathin SiO_2, particularly the barrier height and the effective mass.

Values for the tunneling effective mass in SiO_2 are typically obtained by current–voltage ($I-V$) measurements.[1,3–6] The effective mass is obtained by fitting the $I-V$ measurements to analytic expressions, which are not applicable over a wide bias or oxide thickness range.[3,6] In the present paper, we employ a direct approach to extract the effective mass for tunneling in SiO_2. By characterizing a set of nine different samples with oxide thicknesses in the range 16–35 Å, we are able to observe the exponential dependence of the direct tunneling current on oxide thickness and determine the effective mass and its dependence on device bias. A similar method has been used by Maserjian[7] to investigate the energy-momentum ($E-\kappa$) dispersion relationship in the SiO_2 forbidden gap. However, in Maserjian's paper, there exists significant scatter in the extracted value of the effective mass for low applied bias corresponding to tunneling of electrons with energies up to 1 eV higher than the conduction band of the Si substrate. This low-bias regime is extremely important for the design of scaled MOS devices and resonant tunneling devices. In the present paper, we are able to determine, with a high degree of precision, the effective mass of the tunneling electrons at low applied biases, corresponding to electrons tunneling with energies from 0 to 2 eV above the Si conduction band edge.

An expression that is well suited for the analysis of the thickness dependence of the tunneling current is obtained from Simmons' treatment of a similar problem in metal/insulator/metal structures.[3] From Simmons' paper, the forward tunneling current density for an applied bias smaller than the barrier height can be represented as

$$Jt_{ox}^2=J_0\exp(-2\kappa t_{ox}), \tag{1}$$

where t_{ox} is the oxide thickness, and κ is the magnitude of the electron wave vector in the oxide band gap. The prefactor J_0, derived by Simmons for a symmetric structure with metal electrodes, does not enter into our analysis. Clearly, Simmons' expression for J_0 falls short of a complete description of the current in MOS structures, where one needs to take into account important details such as the different densities of states in the two electrodes as well as the oxide barrier. For precisely this reason, it is incorrect to use Simmons' expression to fit the *amplitude* of the tunneling current in MOS structures, i.e., the entire $I-V$ characteristic for a single sample, to extract the electron effective mass, and may explain the large error bars associated with previous determination of the tunneling effective mass using such a technique.[8] However, in the present work, we analyze the *change* in the tunneling current as a function of thickness, which only requires the portion of Simmons' expression that is derived from basic quantum mechanical considerations, i.e., reflections across a trapezoidal energy barrier, independent of electrode band structure. From the exponential dependence of the product Jt_{ox}^2 on t_{ox}, we can determine κ. The effective mass can then be extracted from κ by using either a parabolic relationship,

$$\kappa=\left(\frac{2m_P^*q[\Phi_B-(V_{ox}/2)]}{\hbar^2}\right)^{1/2}, \tag{2}$$

or a nonparabolic relationship,

$$\kappa=\left[\frac{2m_{NP}^*q[\Phi_B-(V_{ox}/2)]}{\hbar^2}\left(1-\frac{[\Phi_B-(V_{ox}/2)]}{E_G}\right)\right]^{1/2}, \tag{3}$$

where m_P^* is the effective mass in the parabolic $E-\kappa$ approximation Eq. (2), m_{NP}^* is the effective mass in the nonparabolic (two-band) $E-\kappa$ approximation Eq. (3), $q\Phi_B$ is the Al-SiO$_2$ energy barrier height, V_{ox} is the voltage dropped across the oxide, and E_G is the oxide band gap. We assume that the barrier has an average height of $q(\Phi_B - V_{\text{ox}}/2)$, after Simmons.

---

[a] Electronic mail: brar@resbld.csc.ti.com

For our analysis we need to know only two important oxide parameters: (i) The barrier height $q\Phi_B$, which we assume to be 3.17 eV after the work of Maserjian[9] and (ii) the thickness of the oxide t_{ox}, which we determine by both capacitance–voltage ($C-V$) and ellipsometric measurements. The value for the oxide voltage may be assumed to be equal to the applied voltage, or may be determined more accurately by including the surface band bending, which in turn is obtained from the $C-V$ data after Berglund.[10] The difference in the extracted value of the effective mass is less than 3% between the two approaches, and therefore, to keep our analysis simple, we assume that the applied bias is equal to V_{ox}.

The oxides were grown on nominally flat n-type Si(100) wafers with $\rho = 0.01-0.02\ \Omega$ cm. A thermal cleaning process was used for *in situ* surface preparation under ultrahigh vacuum (UHV) conditions with a base pressure of 7×10^{-11} Torr. The oxides were grown in a 400 Torr dry O$_2$ ambient (samples were transferred under UHV). Diodes were fabricated with electron-beam evaporated Al contacts, and $I-V$ and $C-V$ measurements were performed on devices with areas spanning four orders of magnitude in current (15, 150, and 1500 μm diameter). The measured quantities were very reproducible and scaled linearly with the device area, thereby showing excellent uniformity of the oxide layers.

Capacitance–voltage measurements were performed on an HP 4275 LCR meter at 10 kHz. Figure 1 shows the capacitance of the nine samples as a function of applied bias from 0 to 2 V. For our bias polarity, an applied positive voltage to the top Al electrode results in an accumulation in the Si. Capacitance–voltage data through the full two-volt range are obtained for all but the thinnest oxide sample, which is measured only to 0.3 V, limited by the large direct tunneling current through the 16 Å oxide. We determine the oxide capacitance from the $C-V$ measurements in the degenerate accumulation regime, after a method described by Maserjian *et al.*[11] Assuming a bulk dielectric constant for the SiO$_2$ of $\epsilon_{\text{ox}} = 3.9$, we can determine the oxide thickness t_{ox} from the oxide capacitance. For the sample with the thinnest oxide where the $C-V$ data are limited, we estimate a t_{ox} value of 16.5 Å from the zero-bias capacitance, a value which agrees well with our analysis of the tunneling current measured in these structures, and also correlates with the ellipsometric measurement. A comparison of the oxide thicknesses as determined from spectroscopic ellipsometry and $C-V$ measurements is shown in the inset of Fig. 1. We find that the oxide thickness obtained from the $C-V$ data is approximately 3–5 Å larger than the ellipsometric value. This discrepancy is not unusual and may be attributed to the electron wave function in the Si that is peaked away from the oxide barrier, thereby increasing the effective thickness of the SiO$_2$ layer as measured by $C-V$.[12] For our analysis, we use the thickness determined from the $C-V$ data, because both the $C-V$ and $I-V$ measurements are performed on the same device. This provides a better local value for the oxide thickness compared to the ellipsometric value. An advantage of analyzing the *change* in the tunneling current as function of thickness is that a systematic error of 5 Å in the oxide thickness for our samples would change the extracted value of κ by only 2%.

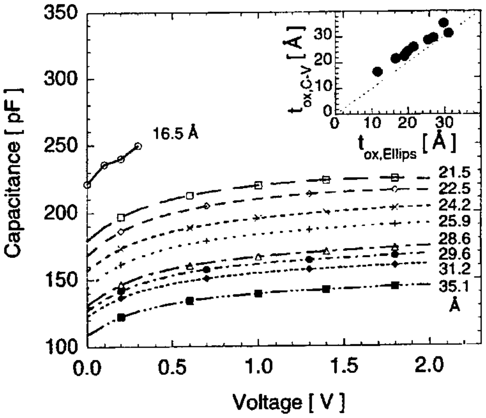


FIG. 1. Capacitance–voltage ($C-V$) measurements in the accumulation region for the nine samples studied. The capacitance for the thinnest oxide cannot be measured above about 0.3 V because of large tunneling currents. The numbers to the right of the curves represent the oxide thickness, which is determined by $C-V$ after the procedure used by Maserjian *et al.* (see Ref. 11) for all but the thinnest oxide. The inset shows a comparison of the oxide thicknesses as measured by $C-V$ and ellipsometry.

Current–voltage measurements were performed using an HP4155 semiconductor parameter analyzer. The current measured in the thickest oxide sample in the low voltage (<0.7 V) range is limited by the sensitivity of the instrument. Figure 2 shows the $I-V$ data in the direct tunneling regime for all nine samples measured on a diode with a diameter of 150 μm. The direct tunneling current is a very sensitive function of the oxide thickness, and is observed to increase over seven orders of magnitude for a decrease in the oxide thickness from 30 to 16 Å. Note that the curves for all the samples are "parallel," indicating that the samples are identical in their electron transport properties. The key point is that the major difference in the $I-V$ curves is in the magnitude of the tunneling current, which is a direct consequence of the different oxide layer thicknesses.

An interesting qualitative aspect of the data in Figs. 1 and 2 is the "correspondence" between the $I-V$ and $C-V$ measurements. From the relative magnitudes, one can clearly identify the $C-V$ curve that corresponds to the $I-V$ curve for the same sample. Inasmuch as the $C-V$ data are used to extract a t_{ox} value for the oxide, such a correspondence suggests that we may also be able to infer a thickness directly from the magnitude of the tunneling currents measured in ultrathin oxides. In the final analysis, the tunneling current is the quantity that interests us most, and is very useful for determining thicknesses in the regime, where the $C-V$ measurements are unreliable as a result of the large tunneling currents.

Figure 3 is a plot of the product Jt_{ox}^2 as a function of t_{ox} at different bias points for eight of the nine samples. The data for the thickest oxide are not included because its tunneling current is too low to measure accurately below 0.7 V.

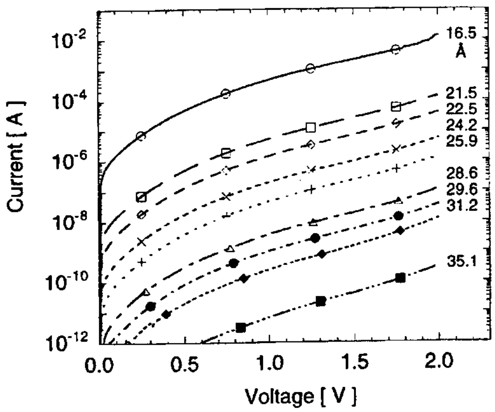


FIG. 2. Oxide thickness dependence of the current–voltage characteristics for nine Al/SiO$_2$/n-Si MOS structures with a diameter of 150 μm. Numbers to the right of the curves are the oxide thickness in Angstroms.

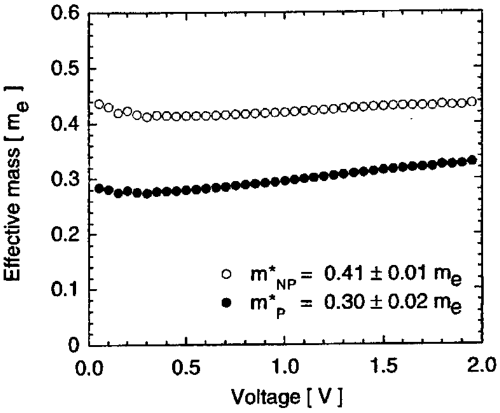


FIG. 4. The tunneling effective mass for the electron in the SiO$_2$ layer extracted from the fit shown in Fig. 3 assuming both a parabolic $E-k$ (m_P^*) and a nonparabolic $E-\kappa$ dispersion (m_{NP}^*) in the oxide, see Eqs. (2) and (3).

The currents for the eight samples fit an exponential dependence on thickness over six orders of magnitude, in excellent agreement with the tunneling current expression in Eq. (1).

Assuming a thick-oxide barrier height of 3.17 eV[9] and using Eqs. (1)–(3) we can extract the effective mass of the tunneling electron at each bias point. Figure 4 is a plot of the electron effective mass in the SiO$_2$ extracted from the thickness dependence of the tunneling currents at various bias points in the direct tunneling regime, using both the parabolic $E-\kappa$ dispersion [Eq. (2); m_P^*], and the nonparabolic dispersion [Eq. (3); m_{NP}^*]. We obtain a value of $m_P^*=0.30 \pm 0.02 m_e$, and $m_{NP}^*=0.42 \pm 0.01 m_e$, over the bias range measured. It is important to note that the effective mass is approximately constant over the bias range studied. The parabolic value is in good agreement with the mass obtained by fits of the tunneling $I-V$ characteristic by Depas *et al.*[5] (0.28–0.31m_e), J. Maserjian[9] (0.27±0.05m_e), and Hiroshima *et al.*[3] (0.33±0.08m_e). The mass obtained from assuming a nonparabolic dispersion is in excellent agreement with Maserjian's value of 0.42m_e for higher bias voltages.[7] Clearly, care must be taken in modeling the direct tunneling current to use the appropriate mass for the dispersion relationship assumed in the oxide forbidden gap. The fact that m_P^* and m_{NP}^* extracted using this simple procedure are essentially independent of the applied bias, and that the data are obtained from $I-V$ curves measured from eight different samples, with t_{ox} ranging from 16.5 to 35 Å, gives better confidence in the value obtained for the effective mass of tunneling electrons in the SiO$_2$ barrier.

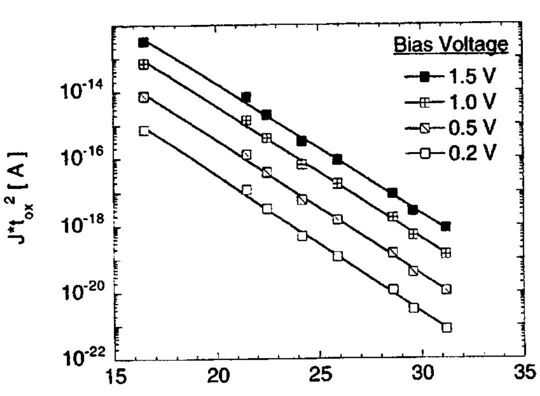


FIG. 3. Exponential dependence of the direct tunneling current on the thickness t_{ox} of the SiO$_2$ insulating layer at four different bias voltages for eight Al/SiO$_2$/n-Si MOS structures. The lines represent an exponential fit to the data spanning six orders of magnitude.

A closer inspection of the dependence of the effective mass on bias in Fig. 4 shows that m_P^* and m_{NP}^* both increase slightly with applied bias (about 3% from 0 to 2 V). A more accurate analysis that includes the effect of band bending in the Si accumulation layer corrects the extracted value of the effective mass in the desired manner, making the data in Fig. 4 flatter, independent of applied bias. However, such a calculation does not add any new physics to our analysis, and only changes the value of the effective mass by less than 3% at 2 V. From the fit to our data, we find that the nonparabolic $E-\kappa$ dispersion provides a more accurate description of the bands in the oxide forbidden gap. This is consistent with the observations of Maserjian for tunneling electrons with energies closer to the conduction band edge of the SiO$_2$.[7]

The authors gratefully acknowledge M. Anthony, G. Brown, R. Lake, T. Moise, and J. P. A. van der Wagt for several stimulating discussions, and they thank M. Howell for assistance in sample preparation. This research has been supported by DARPA Contract No. F49620-96-C-0006.

[1] R. Tsu, A. Filios, C. Lofgren, D. Cahill, J. Vannostrand, and C. G. Wang, Solid-State Electron. (to be published).
[2] A. C. Seabaugh, C.-C. Cho, R. M. Steinhoff, T. S. Moise, K. H. Park, and Y. Okuno, Future Electron Devices J. (to be published).
[3] M. Hiroshima, T. Yasaka, S. Miyazaki, and M. Hirose, Jpn. J. Appl. Phys. **33**, 395 (1994).
[4] S. Zafar, K. A. Conrad, Q. Liu, and E. A. Irene, Appl. Phys. Lett. **67**, 1031 (1995).
[5] M. Depas, R. L. V. Meirhaeghe, W. H. Laflere, and F. Cardon, Solid-State Electron. **37**, 433 (1994).
[6] A. Faigon and F. Campabadal, Solid-State Electron. **39**, 251 (1996).
[7] J. Maserjian, J. Vac. Sci. Technol. **11**, 996 (1974).
[8] G. Simmons, J. Appl. Phys. **34**, 1793 (1963).
[9] J. Maserjian, in *The Physics and Chemistry of* SiO$_2$ *and the Si/*SiO$_2$ *Interface*, edited by C. R. Helms and B. E. Deal (Plenum, New York, 1988).
[10] C. N. Berglund, IEEE Trans. Electron Devices **ED-13**, 701 (1966).
[11] J. Maserjian, G. Petersson, and C. Svensson, Solid-State Electron. **17**, 335 (1974).
[12] R. Rios and N. D. Arora, in IEDM, 1994, p. 613.

Direct extraction of the electron tunneling effective mass in ultrathin SiO_2

B. Brar, G. D. Wilk,* and A. C. Seabaugh

*System Components Laboratory and *Materials Science Laboratory*
Corporate R&D, Texas Instruments, Dallas, TX 75243

ABSTRACT

Electron transport in ultrathin (t_{ox} < 40 Å) Al/SiO_2/n-Si single-barrier structures is dominated by direct tunneling of electrons across the SiO_2 barrier. By analyzing the tunneling currents as a function of the SiO_2 layer thickness for a comprehensive set of otherwise identical samples, we are able to extract an effective mass for the tunneling electron in the SiO_2 layer. Oxide films 16 to 35 Å thick were thermally grown *in situ* in a dry oxygen ambient. The oxide thicknesses were determined by capacitance-voltage measurements and by spectroscopic ellipsometry. The tunneling effective mass was extracted from the thickness dependence of the direct tunneling current between an applied voltage of 0 and 2 V, a bias range that has not been previously explored. Employing both a parabolic and a non-parabolic assumption of the *E-k* relationship in the oxide forbidden gap, we found the SiO_2 electron mass to be $m^*_P = 0.30 \pm 0.02\ m_e$, and $m^*_{NP} = 0.41 \pm 0.01\ m_e$ respectively, independent of bias. Because this method is based on a large sample set, the uncertainty in the mass determination is significantly reduced over prior current-voltage fitting methods. Current-voltage measurements on preliminary Al/SiO_2/Si/SiO_2/n-Si double-barrier structures are also presented.

With the scaling of metal/oxide/semiconductor (MOS) transistors to smaller dimensions, and the realization of silicon-based quantum devices, there has been a renewed interest in the growth and characterization of ultrathin SiO_2 barriers. Oxide thicknesses in the range 50-100 Å are commonplace today in manufacturing, and research into sub-50 Å oxides has intensified. Below 50 Å, the quantum mechanical tunneling current becomes a limiting leakage component in MOS technology, and an enabling element in resonant tunneling and light emitting silicon devices[1,2]. Silicon-based resonant tunneling structures offer the combined advantage of increased density as well as higher functionality; two properties that are important ingredients in bridging the gap between micro- and nano-electronics. In order to fully exploit the promise of silicon quantum devices, it is important to obtain a fundamental understanding of the nature of the electron transport across the SiO_2 barriers, and perform reliable measurements of the basic tunneling properties of ultrathin SiO_2, particularly the energy barrier height and the effective mass.

Values for the tunneling effective mass in SiO_2 are typically obtained by current-voltage measurements[1,3-6]. The effective mass is usually obtained by fitting the *I-V* measurements to analytic expressions, which are not applicable over a wide bias or oxide thickness range[3,6]. In our present paper, we employ a direct approach to extract the effective mass for tunneling in SiO_2. By characterizing a set of nine different samples with oxide thicknesses in the range 16 to 35 Å, we are able to observe the exponential dependence of the direct tunneling current on oxide thickness and determine the effective mass and its dependence on device bias. A similar method has been used by Maserjian[7]. to investigate the energy-momentum (E-κ) dispersion relationship in the SiO_2 forbidden gap. However, in Maserjian's paper, there exists significant scatter in the extracted value of the effective mass for low applied bias corresponding to tunneling of electrons with energies up to 1 eV higher than the conduction band of the Si substrate. This low-bias regime is extremely important for the design of scaled MOS devices and resonant tunneling devices. In the present paper, we are able to determine, with a high degree of precision, the effective mass of the tunneling electrons at low applied biases, corresponding to electrons tunneling with energies from 0 to 2 eV above the Si conduction band edge.

An expression that is well suited for the analysis of the thickness dependence of the tunneling current is obtained from Simmons' treatment of a similar problem in metal/insulator/metal structures, in which the author used the WKB approximation to obtain an analytic expression for the tunneling current[3]. From Simmons' paper, the forward tunneling current density for a metal/insulator/metal structure with an applied bias smaller than the barrier height can be represented as:

$$J t_{ox}^2 = J_0 \exp\left[-2 k t_{ox}\right] \qquad (1)$$

where t_{ox} is the oxide thickness, and k is the magnitude of the electron wave vector in the oxide bandgap. The pre-factor J_0, derived by Simmons for a symmetric structure with metal electrodes, does not enter into our analysis. Clearly, Simmons' expression for J_0 falls short of a complete description of the current in MOS structures, where one needs to take into account important details such as the different densities of states in the two electrodes as well as the oxide barrier. For precisely this reason, it is incorrect to use Simmons' expression to fit the *amplitude* of the tunneling current in MOS structures, *i.e.* the entire *I-V* characteristic for a single sample, to extract the electron effective mass, and may explain the large error bars associated with previous determination of the tunneling effective mass using such a technique[8]. However, in the present work, we analyze the *change* in the tunneling current as a function of thickness, which only requires the portion of Simmons' expression that is derived from basic quantum mechanical considerations, *i.e.* reflections across a trapezoidal energy barrier, independent of electrode band structure. From the exponential dependence of the product Jt_{ox}^2 on t_{ox}, we can determine k . The effective mass can then be extracted from k by using either a parabolic relationship,

$$k = \left(\frac{2\, m_{P}^{*}\, q \left(F_B - \frac{V_{ox}}{2} \right)}{\hbar^2} \right)^{\frac{1}{2}} \qquad (2)$$

or a non-parabolic relationship,

$$k = \left(\frac{2\, m_{NP}^{*}\, q \left(F_B - \frac{V_{ox}}{2} \right)}{\hbar^2} \left(1 - \frac{\left(F_B - \frac{V_{ox}}{2} \right)}{E_G} \right) \right)^{\frac{1}{2}} \qquad (3)$$

where m^*_P is the effective mass in the parabolic *E-k* approximation Eq. (2), m^*_{NP} is the effective mass in the non-parabolic (two-band) *E-k* approximation Eq. (3), qF_B is the Al-SiO$_2$ energy barrier height, V_{ox} is the voltage dropped across the oxide, and E_G is the oxide bandgap. We assume that the barrier has an average height of $q(F_B - V_{ox}/2)$, after Simmons.

For our analysis we need to know only two important parameters associated with the oxide layers: (*i*) The barrier height $q\Phi_B$, which we assume to be 3.17 eV after the work of Maserjian *et al.*[9]. and (*ii*) the thickness of the oxide t_{ox}, which we determine by both capacitance-voltage (*C-V*) and ellipsometric measurements. The value for the oxide voltage may be assumed to be equal to the applied voltage, or may be determined more accurately by including the surface band-bending, which in turn is obtained from the *C-V* data after Berglund[10]. The difference in the extracted value of the effective mass is less than 3% between the two approaches, and therefore, to keep our analysis simple, we assume that the applied bias is equal to V_{ox}.

A detailed description of the oxide growth process may be found in the paper by Wilk *et al.*[11]. No post-oxidation anneals were done on these samples. Diodes were fabricated with electron-beam evaporated Al contacts, and *I-V* and *C-V* measurements were performed on devices with areas spanning four orders of magnitude in current (15, 150, and 1500 μm diameter). The measured quantities were found to be very reproducible and scaled linearly with the device area, thereby showing excellent uniformity of the oxide layers.

Capacitance-voltage measurements were performed on an HP 4275 LCR meter at a measurement frequency of 10 kHz[11]. We determine the oxide capacitance from the *C-V* measurements in the degenerate accumulation regime, after a method described by Maserjian *et al.*[12].. Assuming a bulk dielectric constant for the SiO$_2$ of ε_{ox} = 3.9, we can determine the oxide thickness t_{ox} from the oxide capacitance. For the sample with the thinnest oxide where the *C-V* data is limited, we estimate a t_{ox} value of 16.5 Å from the zero-bias capacitance (measurable for all samples), a value which we show agrees well with our analysis of the tunneling current measured in these structures. A comparison of the oxide thicknesses as determined from spectroscopic ellipsometry and *C-V* measurements shows that the oxide thickness obtained from the *C-V* data is approximately 3 to 5 Å larger than the ellipsometric value. This discrepancy is not unusual and may be attributed to the electron wave function in the Si (at the lower oxide interface) that is peaked away from the oxide barrier, thereby increasing the effective thickness of the SiO$_2$ layer as measured by *C-V*[13]. For our analysis we use the thickness determined from the *C-V* data, because both the *C-V* and *I-V* measurements are performed on the same device. This provides a better local value for the oxide thickness compared to the ellipsometric value. An advantage of analyzing the *change* in the tunneling current as function of thickness is that a systematic error of 5 Å in the oxide thickness for our samples would change the extracted value of κ by only 2%.

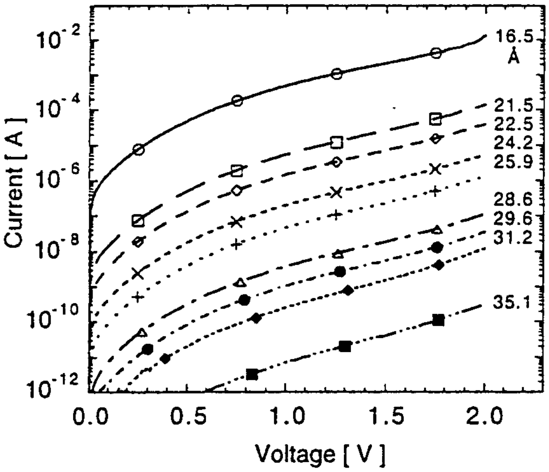


Fig. 1. Oxide thickness dependence of the current-voltage measurements on nine Al/SiO$_2$/n-Si MOS structures with a diameter of 150 μm. The bias is applied to the top Al electrode. The numbers to the right of the curves represent the oxide thickness in Angstroms.

Current-voltage measurements were performed using an HP4155 semiconductor parameter analyzer. The current measured in the thickest oxide sample in the low voltage (< 0.7 V) range is limited by the sensitivity of the instrument. Figure 1 shows the *I-V* data in the direct tunneling regime for all nine samples measured on a diode with a diameter of 150 µm. Fowler-Nordheim tunneling in these structures is not observed until the voltage is increased above 3 V. The direct tunneling current is a very sensitive function of the oxide thickness, and is observed to increase over seven orders of magnitude for a decrease in the oxide thickness from 30 to 16 Å. Note that the curves for all the samples are "parallel," indicating that the samples are identical in their electron transport properties. The key point is that the major difference in the *I-V* curves is in the magnitude of the tunneling current, which is a direct consequence of the different oxide layer thicknesses.

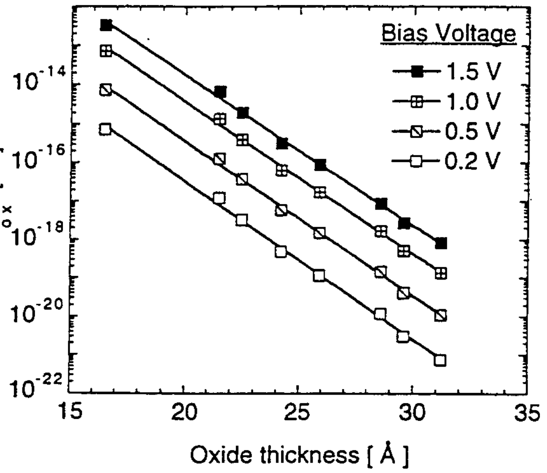


Fig. 2. Exponential dependence of the direct tunneling current on the thickness t_{ox} of the SiO_2 insulating layer at four different bias voltages for nine $Al/SiO_2/n$-Si MOS structures. The lines represent an exponential fit to the data spanning six orders of magnitude. The tunneling effective mass for the electron in the SiO_2 is extracted from the exponential fit using Eqs. (1), (2) and (3).

Figure 2 is a plot of the product Jt_{ox}^2 as a function of t_{ox} at different bias points for eight of the nine samples. The data for the thickest oxide is not included because its tunneling current is too low to measure accurately below 0.7 V. The currents for the eight samples fit an exponential dependence on thickness over six orders of magnitude, in excellent agreement with the tunneling current expression in Eq. (1).

Assuming a thick-oxide barrier height of 3.17 eV[9] and using Eqs. (1), (2) and (3), we can extract the effective mass of the tunneling electron at each bias point. Figure 3 is a plot of the electron effective mass in the SiO_2 extracted from the thickness-dependence of the tunneling currents at various bias points in the direct tunneling regime, using both the parabolic *E-κ* dispersion (Eq. (2); m^*_P), and the non-parabolic dispersion (Eq. (3); m^*_{NP}). We obtain a value of $m^*_P = 0.30 \pm 0.02$ m_e, and $m^*_{NP} = 0.42 \pm 0.01$ m_e, over the bias range measured. It is important to note that the effective mass is approximately constant over the bias range studied. The parabolic value is in good agreement with the mass obtained by fits of the tunneling *I-V* characteristic by Depas *et al.*[5] (0.28 to 0.31 m_e), Maserjian *et al.*[9] (0.27 ± 0.05 m_e), and Hiroshima *et al.*[3] (0.33 ± 0.08 m_e). The mass obtained from assuming a non-parabolic dispersion is in excellent agreement with Maserjian's value of 0.42 m_e for higher bias voltages[7]. Clearly, care must be taken in modeling the direct tunneling current to use the appropriate mass for the dispersion relationship assumed in the oxide forbidden gap. The fact that m^*_P and m^*_{NP} extracted using this simple procedure are essentially independent of the applied bias, and that the data are obtained from *I-V* curves measured from eight different samples, with t_{ox} ranging from 16.5 to 35 Å, gives better confidence in the value obtained for the effective mass of tunneling electrons in the SiO_2 barrier.

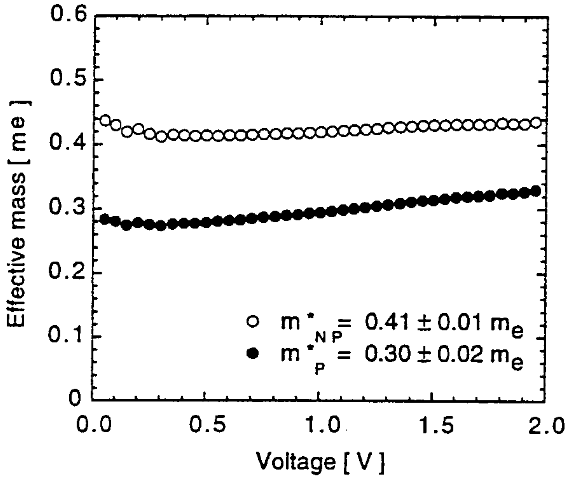


Fig. 3. The tunneling effective mass for the electron in the SiO_2 layer extracted from the fit shown in Fig. 2 assuming both a parabolic E-k (m^*_P) and a non-parabolic *E-κ* dispersion (m^*_{NP}) in the oxide, see Eqs. (2) and (3).

A closer inspection of the dependence of the effective mass on bias in Fig. 3 shows that m^*_P and m^*_{NP} both increase slightly with applied bias (about 3% from 0 to 2V). A more accurate analysis that includes the effect of band-bending in the Si accumulation layer corrects the extracted value of the effective mass in the desired manner, making the data in Fig. 3 flatter, independent of applied bias. However, such a calculation does not add any new physics to our analysis, and only changes the value of the effective mass by less than 3% at 2 V. From the fit to our data, we find that the non-parabolic *E-κ* dispersion provides a more accurate description of the bands in the oxide forbidden gap. This is consistent with the observations of Maserjian for tunneling electrons with energies closer to the conduction band edge of the SiO_2[7].

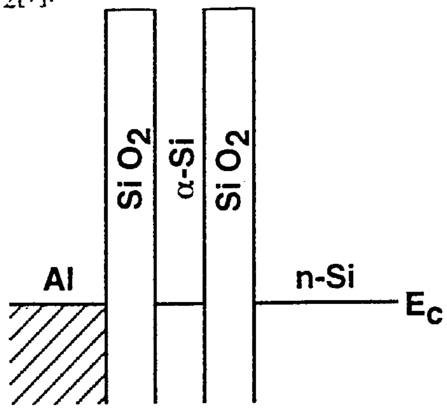


Fig. 4. Energy band diagram of a double-barrier resonant tunneling structure.

In conclusion, we would like to present some preliminary *I-V* data on $Al/SiO_2/Si/SiO_2/n\text{-}Si$ *double*-barrier resonant tunneling diodes. The energy band diagram of the structure is shown in Fig. 4. Such structures have been fabricated in our laboratory using the same technique described in the present paper, which should be compatible with a silicon VLSI process The nominal thicknesses of the constituent layers are approximately 1-2 nm. In this device's simplest incarnation, the Si-layer sandwiched between the SiO_2 layers is amorphous, and the current transport is normal to the semiconductor surface. Room temperature current-voltage (*I-V*) measurements on the double-barrier structure have been performed and are shown in Fig. 5. While there is no evidence of resonant tunneling in this preliminary structure, note that Fowler-Nordheim oscillations are observed in the high bias tunneling regime. By subtracting the non-oscillatory exponential tunneling current from the measured *I-V* characteristic, the periodic nature of these current oscillations can be highlighted, as shown in the inset in Fig. 5. In contrast to conventional Fowler-Nordheim oscillations in single-barrier structures, the oscillatory component in the double-barrier structure shows the presence of at least two frequencies, presumably associated with reflections from both the oxide barriers in the structure. Of course, the Fowler-Nordheim current oscillations are too weak to be useful for practical device applications, but they are an important preliminary result as they clearly indicate that the electrons are able to traverse the structure in a phase-coherent manner, a necessary condition for the realization of resonant tunneling devices.

We gratefully acknowledge M. Anthony, G. Brown, R. Lake, T. Moise, and J. P. A. van der Wagt for several stimulating discussions, and we thank M. Howell for assistance in sample preparation. This research has been supported by DARPA contract F49620-96-C-0006.

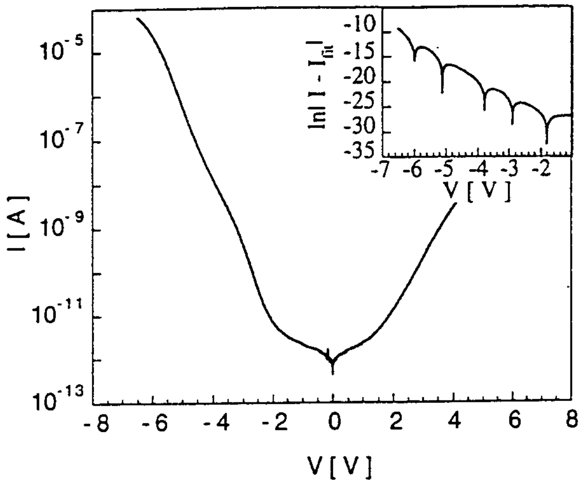


Fig. 5. Current-Voltage characteristics of a double-barrier structure showing Fowler-Nordheim current oscillations indicative of phase-coherent electron transport. The inset shows the oscillatory part of the tunneling current.

References

[1] R. Tsu, A. Filios, C. Lofgren, D. Cahill, J. Vannostrand, and C. G. Wang, Solid State Electronics, to be published (1996).

[2] A. C. Seabaugh, C.-C. Cho, R. M. Steinhoff, T. S. Moise, K. H. Park, and Y. Okuno, Future Electron Devices Journal (1996).

[3] M. Hiroshima, T. Yasaka, S. Miyazaki, and M. Hirose, Japanese Journal of Applied Physics **33,** 395 (1994).

[4] S. Zafar, K. A. Conrad, Q. Liu, and E. A. Irene, Applied Physics Letters **67,** 1031 (1995).

[5] M. Depas, R. L. V. Meirhaeghe, W. H. Laflere, and F. Cardon, Solid-State Electronics **37,** 433 (1994).

[6] A. Faigon and F. Campabadal, Solid-State Electronics **39,** 251 (1996).

[7] J. Maserjian, J. Vac. Sci. Technol. **11,** 996 (1974).

[8] G. Simmons, Journal of Applied Physics **34,** 1793 (1963).

[9] J. Maserjian, in *The Physics and Chemistry of SiO2 and the Si/SiO2 interface*, edited by C. R. Helms and B. E. Deal (Plenum Press, New York, 1988).

[10] C. N. Berglund, IEEE Trans. on Electron Devices **ED-13,** 701 (1966).

[11] G. Wilk, B. Brar, Y. Wei, R. Wallace, and A. Seabaugh, in TI CRTC, Plano, TX, 1996.

[12] J. Maserjian, G. Petersson, and C. Svensson, Solid-State Electronics **17,** 335 (1974).

[13] R. Rios and N. D. Arora, in IEDM, 1994, p. 613.

Resonant Tunneling in Disordered Materials such as $SiO_2/Si/SiO_2$

R. Lake, B. Brar, *G. D. Wilk, A. Seabaugh, and G. Klimeck

Raytheon TI Systems, Box 655936, MS 134, Dallas, TX 75265
*Texas Instruments, Box 655936, MS 147, Dallas, TX 75265

Abstract. We have analyzed the effect of disorder in both the well and barriers of a resonant tunneling diode (RTD). If the disorder is limited solely to the barriers, a good peak-to-valley ratio (PVR) is expected. We describe a general guideline relating the PVR to the bulk mobility and effective mass of the well material of an RTD. We compare the effects of correlated versus uncorrelated disorder on the valley current. We discuss why interband tunnel devices such as the Esaki diode are more robust than RTDs in the presence of disorder.

1. Well versus Barrier Disorder

There have been a number of studies of Si / SiO_2 multilayer structures to look for quantization and resonant effects [1,2]. While the interface can be made smooth, the entire multi-layer structure is amorphous. Core level x-ray spectroscopy gives unambiguous evidence that quantization does occur, and that it follows the standard inverse square relationship to the well width [1]. However, there have been no strong demonstrations of resonant tunneling and negative differential resistance (NDR) in such systems. The observation of state-quantization with core-level-spectroscopy requires only conservation of total energy. State quantization is a necessary but not sufficient condition to observe NDR. Observation of NDR in an RTD requires conservation of both total energy and transverse momentum. The random potential resulting from the noncrystalline nature of the Si/SiO_2 material breaks the translational periodicity and, thus, the transverse momentum conservation required for high peak-to-valley ratios in RTDs.

We simulate the effect of the disorder by creating a mathematical model which mimics the macroscopic effects of amorphous disorder. The model gives the same momentum independence to the scattering since the correlation length of amorphous disorder is approximately the lattice constant, and the model gives the same bulk mobility. We begin with a single-band, tight-binding Hamiltonian which parameterizes the Si-Si (or for III-Vs, the cation-anion) basis matrix elements into a single number, the site energy. Then we assign a random component, δV, to the site energy. The random component has a Gaussian distribution with a mean of 0 and a variance of σ. The random component is uncorrelated between any two sites. The only free parameter in the model is σ. To relate σ to a physical quantity, we use the fact that the mobility resulting from this disorder is

$$\mu = \frac{4e\pi^{1/2}\hbar^4}{3m_c m_d^{3/2}\left(k_B T\right)^{1/2}\Omega\sigma^2} \tag{1}$$

where e is the magnitude of the electron charge, Ω is the volume of the primitive cell, m_c is the conductivity effective mass and m_d is the density of states effective mass. The disorder is treated in the self-consistent Born approximation [3].

In Fig. (1), we choose the tight-binding parameters corresponding to an effective mass of 0.3 for the SiO_2 [4] and 0.5 for the a-Si [1]. In Fig. (1a) the disorder is restricted to the well region. We find that the NDR is destroyed when the mobility of the a-Si in the well is less than or equal to 250 cm^2/Vs. If we take the same structure and restrict the same disorder to the barrier layers, there is almost no effect on the I-V and the PVR as is shown in Fig. (1b).

2. General Guideline Relating the PVR to the Bulk Mobility

When attempting to build an RTD in a new material system that is disordered or even amorphous, we would like to have some experimental measure which will predict whether the RTD will have a useful

CCC Code 0-7803-3883-9/98/$10.00 © 1998 IEEE

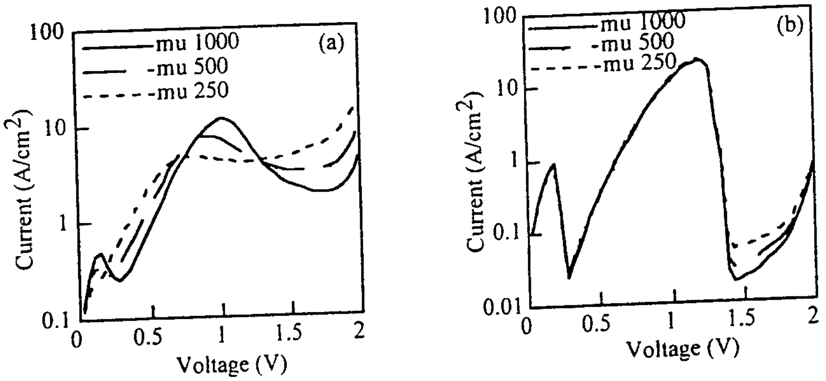


Figure 1. I-V of an amorphous $SiO_2/Si/SiO_2$ RTD with dimensions 1.35/2.97/1.35 (nm) and 10^{18} cm^{-3} doped leads. (a) The disorder is restricted to the well region. (b) The disorder is restricted to the barrier region.

peak-to-valley ratio. We consider the case in which (a) both the mobility of the bulk material and the valley current of the RTD are determined by the static disorder scattering resulting from substitutional disorder (alloys), geometric disorder (amorphous and poly-crystalline materials), impurities and dopants, and (b) the bulk material and the thin layer have the same microscopic structure. Under such circumstances, both analytical [5] and numerical calculations show that the PVR is related to the bulk, room-temperature mobility by

$$\mathrm{PVR} \propto \tau = \mu m_c / e. \qquad (2)$$

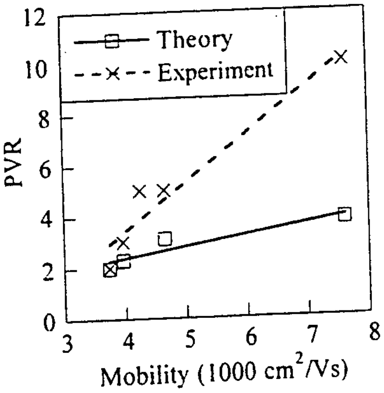


Fig. 2. Experimental and theoretical PVR vs. mobility for an InGaAs well.

Doping provides an experimental knob on the disorder and mobility. In a set of experiments, we increased the doping throughout an AlAs/InGaAs/AlAs based RTD, measured the PVR, and measured the mobility of the well material by growth of bulk InGaAs epi-layers characterized by resistivity and Hall measurements. We also numerically calculated the PVR of the InGaAs RTD as a function of mobility. The experimental and numerical results are shown in Fig. (2). Except at low mobility, the numerical calculations tend to underestimate the PVR. Therefore, we use the experimental data combined with relation (2) to make predictions for other materials.

We propose a general guide relating PVR to the bulk mobility and conductivity effective mass illustrated in Fig. (3). The PVR vs. mobility curve for the InGaAs RTD is replotted. The other curves are obtained by scaling the InGaAs mobility by m_{InGaAs}/m_{new}. The scaling law suggests that to obtain a PVR of 5 in a-Si with $m_c = 0.5\ m_0$ will require a mobility of 400 cm^2/Vs, well above values found for a-Si.

3. Correlated Versus Uncorrelated Disorder

In an ideal epitaxial structure, the periodic crystal potential is undisturbed in the transverse plane, therefore transverse crystal momentum is conserved, and the process illustrated in Fig. (4) is not

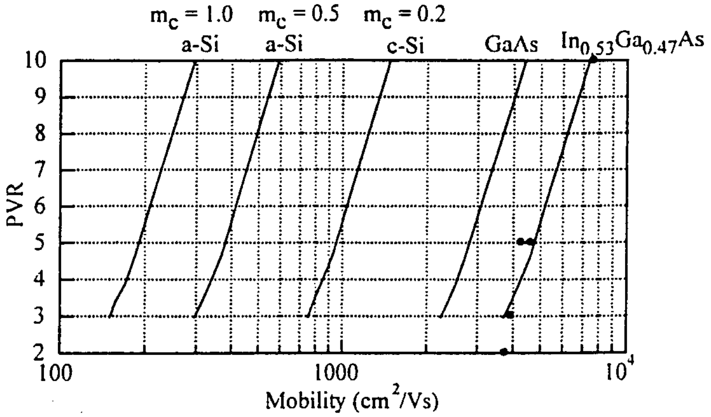


Figure 3. General guide relating PVR to the bulk mobility and conductivity effective mass of the well material.

allowed. In the presence of disorder, the incident electron can pick up a Fourier component of the disorder potential and scatter into a resonant transverse momentum state k_t^r. This is the process by which elastic scattering from disorder contributes to the valley current. The momentum coupling which determines the amount of momentum that an electron can obtain from the disorder is given by the Fourier transform of the random potential autocorrelation function.

When the disorder is uncorrelated $\langle V(\boldsymbol{R})V(\boldsymbol{R}')\rangle \propto U\delta(\boldsymbol{R}-\boldsymbol{R}')$, and the momentum coupling is a constant U. In this case, an incident electron can pick up any transverse momentum that it needs to get into the resonant state, and the scattering component of the valley current is nearly constant, independent of bias as illustrated schematically in Fig. (5a).

If the disorder is correlated, the inverse correlation length provides a cutoff to the momentum transfer. For example with Gaussian correlation, $\langle V(\boldsymbol{R})V(\boldsymbol{R}')\rangle \propto e^{-|\boldsymbol{R}-\boldsymbol{R}'|^2/\Lambda^2}$ where Λ is the correlation length, and the momentum coupling falls off as $|U_q|^2 \propto \Lambda^3\pi^{3/2}e^{-q^2\Lambda^2/4}$. In the valley current region, as the bias is increased, an incident electron must pick up more transverse momentum to scatter into the resonant subband. The amount of momentum that the electron can obtain is governed by $|U_q|^2$. Therefore, for correlated disorder, the scattering component of the valley current falls off with bias as shown schematically in Fig. (5b).

We have numerically explored the effect of the correlation length for disorder limited to the interface layers of GaAs / AlAs RTDs for both exponentially and Gaussian correlated disorder [6]. Figs. (3) and (4) of reference [6] show that as the correlation length increases, the slope of the valley current region becomes more negative and the scattering assisted current decreases with bias. These results indicate that a polycrystalline well with large enough crystal sizes may be sufficient for observing NDR.

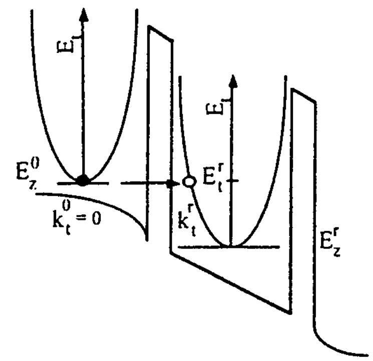


Figure 4. The parabolas represent the transverse kinetic energy. An off-resonant incident electron with longitudinal energy E_z^0 and zero transverse kinetic energy can elastically scatter into the resonant state by trading off longitudinal energy for transverse kinetic energy.

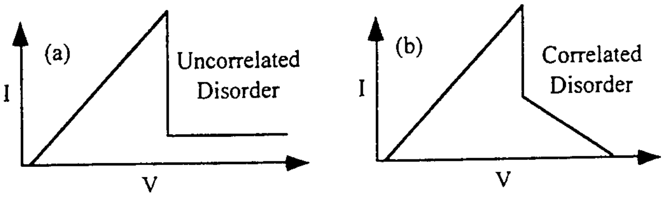


Figure 5. (a) The contribution to the valley current from uncorrelated disorder scattering is essentially flat, independent of bias. (b) The contribution to the valley current from correlated disorder scattering falls off with bias.

Interband Tunnel Diodes and Disorder

Very general reasons suggest that an interband tunnel diode (ITD) such as the Esaki diode [7], the quantum well Esaki diode [8], or the resonant interband tunnel diode [9] should be less affected by disorder than an RTD. A notable difference between an ITD and an RTD is that an ITD requires one less conservation law to exhibit NDR. RTDs require conservation of both total energy and transverse momentum whereas ITDs require only conservation of total energy. For an RTD, Fig. 4 shows that there is always a high transverse-energy state in the well into which an incident electron can scatter while conserving total energy. For an ITD biased in the valley current region, there are no valence states into which an electron can elastically scatter (see for example Figs. (4), (6), and (8) of reference [10]). Since disorder scattering is elastic, the ITD should be more resilient than the RTD to the presence of disorder. This is in fact evident for an Esaki diode since the active region is doped to concentrations of 10^{20} cm^{-3}. Such a high doping in the well of an RTD would destroy its NDR.

Summary

Although there is experimental evidence of coherent resonance effects in amorphous Si / SiO$_2$ materials, both numerical calculations and analytical theory do not show promise for good PVR from such materials. Amorphous barriers with a crystalline well, however, appear to be sufficient. The analysis of correlated disorder indicates that polycrystalline wells may be sufficient if the domain sizes are large enough. Interband devices are more robust in the presence of disorder since they require one less conservation law to work.

References

[1] Lu Z, Lockwood D J, and Barlbeau J M 1995 *Nature* 378 258-260
[2] Hirose M, Morita M and Osake Y (1977) *Jpn. J. Appl. Phys.* 16 Suppl. 16-1 561-4
[3] Lake R, Klimeck G, Bowen R C, and Jovanovic D 1997 *J. Appl. Phys.* 81 7845-7869
[4] Brar B, Wilks G D, and Seabaugh A 1996 *Appl. Phys. Lett.* 69 2728-30
[5] Lake R, Klimeck G., Anantram M P, and Datta S 1993 *Phys. Rev. B* 48 15132-7
[6] Lake R, Klimeck G, Bowen R, Fernando C, Moise T, and Kao Y 1996 *Superlatt. Microstruct.* 20 279-285
[7] Esaki L 1958 Phys. Rev. 109 603-4
[8] Tsai H H., Su Y K, Lin H H, Wang R L, Lee . L 1994 *IEEE Electron Device Lett.* 15 357-359
[9] Söderström J R, Chow D H, McGill M C (1989), *Appl. Phys. Lett.* 55 1094-1096
[10] Seabaugh A and Lake R 1997 *Encyclopedia of Applied Physics vol.* 22 (New York:VCH Publishers) 1-24

Electronic properties of an epitaxial SiO_x/Si heterojunction

B. Brar, G. D. Wilk*, and A. C. Seabaugh

Raytheon Systems, Dallas, TX, USA
*Texas Instruments, Dallas, TX, USA

Abstract

Crystalline silicon suboxide, SiO_x ($0 < x < 2$), films with oxygen contents ranging from 10 to 35 atomic percent have been grown on Si(100) substrates. Temperature-dependent current-voltage measurements on Al/SiO_x/Si diodes fabricated from such structures show that the transport across a SiO_x layer with 30 at. % oxygen is characterized by a thermally activated process with an activation energy of 0.5 eV. The bias dependence of the extracted activation energy and the observed current-voltage characteristics indicates that the measured activation energy corresponds to the conduction band-offset between the SiO_x and Si.

Introduction

An insulating layer, lattice-matched to silicon, with a tunable energy bandgap would be a revolutionary material to introduce to the silicon semiconductor industry. Such a material could enable heterostructure devices and bandgap engineering in silicon; a welcome addition to the device designers toolbox to meet the formidable challenge of scaling devices to the nanometer length-scale. Towards this end, recent work has demonstrated that crystalline SiO_x ($0 < x < 2$) suboxide layers may be grown under conditions straddling the phase boundary between SiO_2 oxidation and SiO etching of the substrate.[1-4] In the present paper, we demonstrate that the SiO_x films are sufficiently crystalline that single crystal silicon may be overgrown on such SiO_x films; and that the SiO_x layer acts as an insulating barrier with a conduction band-offset determined to be approximately 0.5 eV with respect to silicon.

Growth and Experiment

SiO_x/Si samples were grown on 4-inch n-type Si(100) wafers with $\rho = 0.01 - 0.02\ \Omega$-cm. Prior to being placed in an ultrahigh vacuum (UHV) chamber, the wafers were etched in a buffered HF solution (2% HF) for 20 seconds to provide an H-terminated surface. The base pressure in the UHV chamber was 7×10^{-11} Torr. After being a degassed at 400°C for several hours, the substrates were heated up to 700°C for 10 minutes to remove the hydrogen from the surface and to obtain a clear 2 x 1 reflection high-energy electron diffraction (RHEED) pattern. The SiO_x films were deposited in the temperature range 680 – 740°C by backfilling the chamber with oxygen to pressures ranging from 7×10^{-6} to 5×10^{-5} Torr. An electron beam evaporator was used to deposit the Si at a rate of about 2 Å/sec. Diodes were fabricated with Al contacts, and temperature-dependent current-voltage (I-V) measurements were performed on devices spanning four orders of magnitude in current (15, 150, and 1500 µm in diameter). The measured quantities were reproducible and scaled linearly with the device area.

Results and Discussion

Figure 1 shows a cross-sectional transmission electron microscope (TEM) image of a 500 Å SiO_x layer with 1500 Å of single crystal Si grown on top of it. While stacking faults can be seen in the overgrown film, it is clearly single crystal. RHEED patterns showing 2 x 1 reconstruction also confirmed the existence of a crystalline Si surface throughout the growth of the Si overlayer. High-resolution TEM (not shown here)

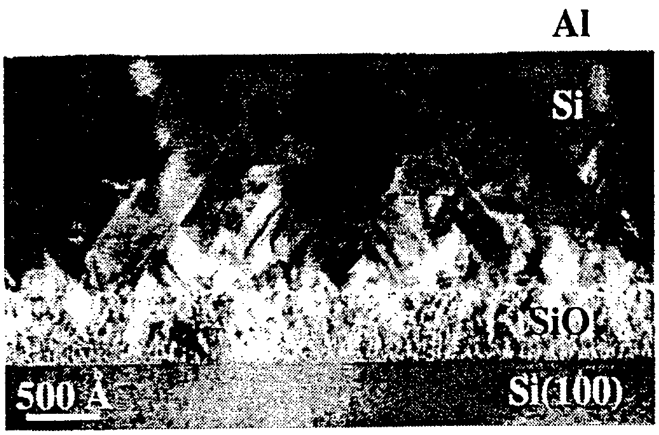


Fig. 1. Cross-sectional TEM of an Al/Si/SiO_x/Si epitaxial structure. The 1500 Å Si layer above the 500 Å thick SiOx layer is single crystal.

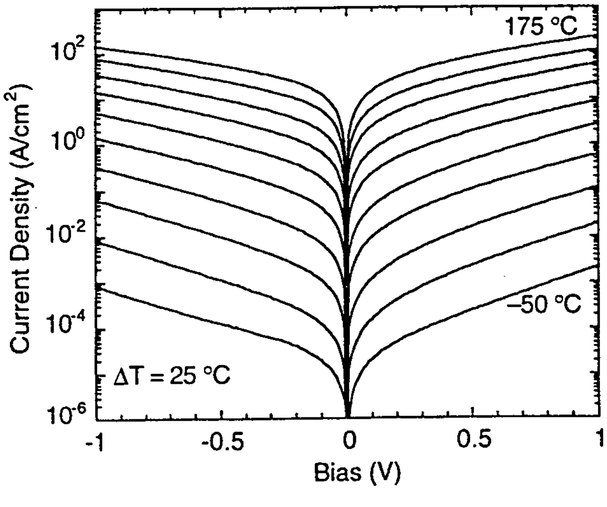


Fig. 2. Current-Voltage (I-V) characteristics of an Al/SiO_x/Si structure at various temperatures. The temperature ranges from –50°C to 175 °C in steps of 25 °C.

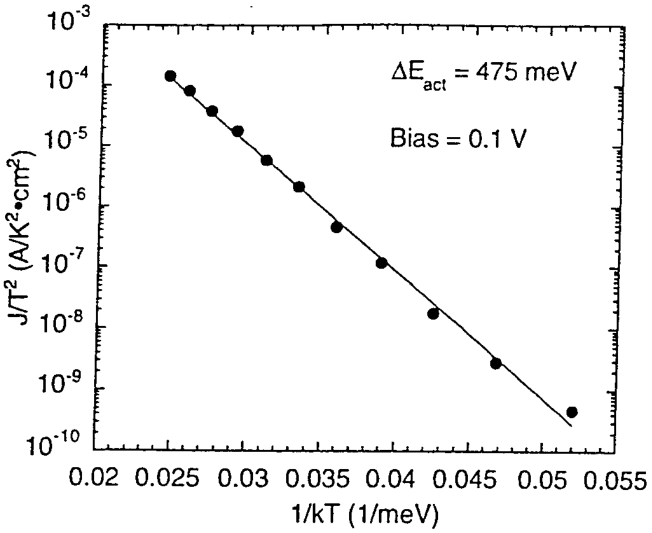


Fig. 3. Arrhenius plot of the current transport across the SiO_x layer. The data is represented by the black dots and the line represents a fit to a thermionic emission model over almost six orders of magnitude.

shows that the layer is inhomogeneous, with neighboring regions of order and disorder, corresponding to phases with different oxygen levels, within the SiO_x film. The regions of order allow single crystal overgrowth of Si at relatively low temperatures of about 700°C. Al was deposited on the Si layer to provide an electrical contact, as well as to provide a marker layer in TEM. Since obtaining single crystal Si overgrowth by depositing directly on a silicon oxide layer is difficult, this demonstrates that crystalline Si can be obtained with epitaxial control on an insulating SiO_x layer. The oxygen levels and degree of crystallinity in the SiO_x films have been measured by Rutherford backscattering spectrometry and ion channeling measurements, respectively, with χ_{min} values as low as 31% for films with 15 at. % oxygen.

Temperature-dependent I-V measurements were performed on an $Al/SiO_x/Si$ sample with 30 at. % oxygen. Figure 2 shows the I-V characteristics for various temperatures ranging from -50 to 175 °C. The I-V characteristics are symmetric about zero bias indicating a symmetric barrier corresponding to the SiO_x layer sandwiched between the top Al contact and the n-type Si substrate. The weak dependence on applied bias is indicative of carrier transport across a heterojunction with a fixed barrier height. The barrier height can be obtained from the activation energy associated with the temperature dependence of the current. Figure 3 is an Arrhenius plot of the temperature dependence of the current at a bias of 0.1 V. The solid circles are the data at various temperatures, and the line represents a fit to a thermionic emission model for the current transport. An excellent fit is obtained over almost six orders of magnitude. An activation energy of 475 meV is obtained from the model.

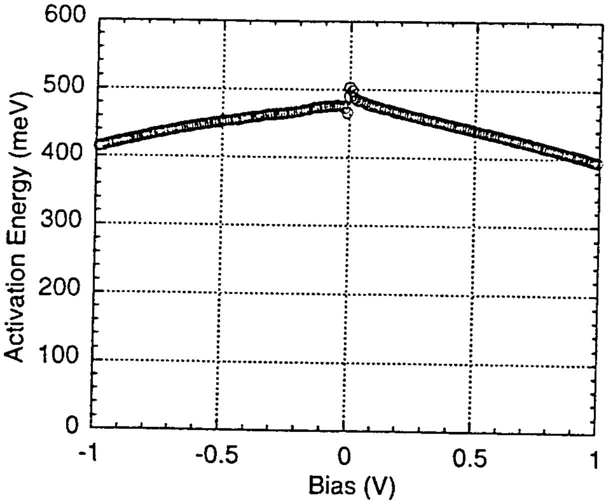


Fig. 4. The extracted activation energy as a function of bias obtained from a thermionic emission model. The weak dependence on applied bias is consistent with the I-V characteristics in Fig. 2, and indicates a fixed heterostructure barrier associated with the SiO_x layer.

To understand the transport mechanism in more detail, we extract the activation energy at each voltage shown in Fig. 2 to study the bias dependence of the activation energy as shown in Fig. 4. The activation energy in Fig. 4 depends very weakly on applied bias, indicative of a fixed heterojunction barrier. The energy barrier to current transport across the SiO_x layer is given by the measured activation energy at approximately zero bias: ≈ 0.5 eV. The slight decrease of the barrier height with applied bias is presumably due to image-force lowering of the barrier and/or tunneling through the tip of the barrier. Note that the bias-dependence of the activation energy in Fig. 4 is also symmetric about zero bias, consistent with the bias-dependence of the current in Fig. 2.

We believe that the simplest explanation of the transport mechanism is electronic conduction across the SiO_x conduction band barrier. The barrier height to both Al and Si is ≈0.5 eV. The equivalence of the barrier height to both Al and Si in the $Al/SiO_x/Si$ structures is not surprising, considering the same behavior is observed for $Al/SiO_2/Si$ and Al/ZnS/Si structures.[5]

Conclusions

The physical and electrical properties of a new crystalline oxide barrier have been determined. The oxide, SiO_x with $x < 2$, yields a conduction band offset of ≈0.5 eV with Si. It is likely that control of the oxygen content can be used to adjust the band-offset and provide a tunable barrier height.

References

[1] T.J. Maloney, J. Vac. Sci. Technol. B **1**, 773 (1983).
[2] M. Hacke, H.L. Bay and S. Mantl, Thin Solid Films **280**, 107 (1996).
[3] M. Tabe, M. Takahashi Y. Sakakibara, Jpn. J. Appl. Phys. **26**, 1830 (1987).
[4] G.D. Wilk, B. Brar, to be published.
[5] B. Brar, R. Steinhoff, A. Seabaugh, X. Zhou, S. Jiang, and W. P. Kirk, Proceedings of the 24th International Symposium on Compound Semiconductors, San Diego (1997.

Appendix 9

Crystalline-Silicon Quantum Well (CQW) Process Traveler
for fabrication of $SiO_2/Si/SiO_2$ resonant tunneling diodes

Clean-up for Rapid Thermal Oxidation (RTO)

- Soak 2 min in $1H_20:1NH_4OH$ (ammonium hydroxide) to remove organics
- Rinse 2 min in DI
- Soak 2 min in $1H_2O:1HCl$
- Rinse 2 min in DI
- Etch off surface oxide in buffered HF [$1HF:100H_20$] 25 s
- Blow dry (no rinse, BHF should roll off the wafer; if it doesn't repeat the previous step and note the time)
- Inspect
- RTO 1040 °C 40 s to grow a 40 Å oxide

Forming Gas Anneal

- Furnace anneal in forming gas at 430 C for 30 minutes

ALIGNMENT MARKER FORMATION BY ETCHING TRENCHES

Shipley 510 RESIST PROCESS – 1.25 μm thickness

Resist Coater Program 1	or	Hand Coat
• Bake 130 °C 30 s - HMDS - Bake		
• Spin 1000 rpm 3 s		• Slow Rise 130 °C
• Dispense Shipley 510 200 rpm 5 s		• Spin HMDS 2500 rpm 20 s
• Spin 2500 rpm 20 s		• Spin Shipley 510 2500 rpm 20 s
• Dry Spin 2500 rpm 15 s		• Hot plate prebake 90 °C 90 s
• Hot plate prebake 90 °C 60 s		

- Expose 200 ms **_reticle - TI0097, Pattern: CQW-MARK1_**

Resist Developer Program 1	or	Hand develop
• Post exposure bake 110 °C 90 s		
• Spin 500 rpm 2 s		
• Prewet rinse 500 rpm 4.5 s		• Post exposure bake 110 °C 90 s
• Spray NMD-W 400 rpm 65 s		• Develop NMD-W 1 min
• DI rinse 1000 rpm 20 s		• Rinse DI 30 s
• Spin 5000 rpm 20 s		

Etch in PlasmaTherm 700 then Evaporate Markers into Silicon Trenches

Etch conditions:	Pressure	25 mTorr	Power	100 W
	CF_4 rate	40 sccm	O_2 rate	3 sccm
	Temperature	35 °C		
	Si Etch Rate	~2 Å/s		

Etch time 10 min. Target etch depth 1200 Å

- Load without delay into evaporator
- Evaporate: 400 Å Ti
  250 Å Pt
- Lift-off
- Soak in warm acetone 2 min
- Soak in warm methanol 2 min
- Blow dry with nitrogen
- Spin rinse, spin dry if necessary
- Inspect
- Step profile in three places across the wafer diameter:

NewMesa - Formation of Micron-To-Submicron Size Resist Islands

- Inspect under microscope, preclean as necessary, record steps taken

Shipley 510 RESIST PROCESS – 1.25 µm thickness

Resist Coater Program 1	or	Hand Coat
• Bake 130 °C 30 s - HMDS - Bake		
• Spin 1000 rpm 3 s		• Slow Rise 130 °C
• Dispense Shipley 510 200 rpm 5 s		• Spin HMDS 2500 rpm 20 s
• Spin 2500 rpm 20 s		• Spin Shipley 510 2500 rpm 20 s
• Dry Spin 2500 rpm 15 s		• Hot plate prebake 90 °C 90 s
• Hot plate prebake 90 °C 60 s		

- Expose 200 ms ***reticle - TI0097-02B, Patterns: CQW - NEWMESA***

Resist Developer Program 1	or	Hand develop
• Post exposure bake 110 °C 90 s		
• Spin 500 rpm 2 s		
• Prewet rinse 500 rpm 4.5 s		• Post exposure bake 110 °C 90 s
• Spray NMD-W 400 rpm 65 s		• Develop NMD-W 1 min
• DI rinse 1000 rpm 20 s		• Rinse DI 30 s
• Spin 5000 rpm 20 s		
• Bake 120 °C 60 s		

- Inspect

Etching to Form SiO_2 Islands

- Prepare $100H_2O$:1HF etch
  Etch rate is approximately 1 Å/s
- Rinse in DI 1 min
- Asher 1 min 8 W in O_2
- Etch in $100H_2O$:1HF to remove SiO_2, 20 s [SiO_2 thickness x 1A/sec + 10 sec]
- Rinse in DI 10 s
- Blow dry with nitrogen
- We expect water to roll off the wafer after the 100:1 etch; does it? Yes No.
- Remove resist in acetone spray
- Soak in warm acetone 5 min
- Soak in warm methanol 5 min
- Rinse in DI - NO SPIN RINSE SPIN DRY because heating removes HF treatment
- Blow dry with nitrogen

Molecular Beam Epitaxy (MBE) - Growth of Silicon Quantum Well

Prior to MBE growth etching, oxide island thickness is 45 Å

- Rinse in DI 1 min
- Etch in $100H_2O$: 1HF 25-35 s to thin oxides to 20-10 Å (etch rate 1 Å/s)
- Load without delay
- Grow 30 Å Si quantum well
- Form second barrier by UVO
- Record growth condition and observations

Second Forming Gas Anneal

Wafers are already clean

- Furnace anneal in forming gas at 430 °C for 30 minutes
  3 min out/30 minutes in/3 minutes out

Thin Aluminum Evaporation

- Inspect.
- Load without delay
- Evaporate 500 Å of Aluminum

E-Beam Emitter

- Spin 950 PMMA (polymethylmethacrylate) 1400 rpm 30 s
- Hot plate bake 180 °C 60 s (This needs to be in a hood)
- Ebeam:
  Circuit: CQW Pattern: EMITTER4 Dose: 320 $\mu C/cm^2$ Mode: 6 Marker: 4

- Develop in APT: 1 MIBK:1 IPA 180 s (Prog 2)
- Post exposure hot plate bake at 100 °C 60 s
- Step profile to get PMMA height _____ Å
- Inspect
- 50 W 30 s oxygen asher
- Load without delay
- Evaporate:

400	Å	Ti
200	Å	Pt
200	Å	Pt
1200	Å	Au
2000	Å	Au

- Tape lift-off
- Remove PMMA in acetone
- Soak in warm acetone 5 min
- Soak in warm methanol 5 min
- Blow dry with N_2
- Inspect
  Step profile Au metal thickness in three places:

Large Emitter Metallization (Contact Print)

AZ1915 RESIST PROCESS – 1.5 µm thickness

Resist Coater Program 3	or	Hand Coat
• Bake 130 °C 30 s - HMDS - Bake		
• Spin 1000 rpm 3 s		• Slow Rise 130 °C
• Dispense AZ1915 400 rpm 5 s		• Spin HMDS 4000 rpm 20 s
• Spin 4000 rpm 20 s		• Spin AZ1915 4000 rpm 20 s
• Dry Spin 2500 rpm 15 s		• Hot plate prebake 120 °C 60 s
• Hot plate prebake 120 °C 60 s		

- Align and expose on contact aligner ***Use mask: CQW-METAL4***

Resist Developer Program 2	or	Hand develop
• No post exposure bake		
• Spin 500 rpm 2 s		• No post exposure bake
• Prewet rinse 400 rpm 4.5 s		• Develop 1AZ:1H_2O 4 min
• Spray 1AZ:1H_2O 400 rpm 90 s		• Rinse DI 30 s
• DI rinse 1000 rpm 20 s		
• Spin 5000 rpm 20 s		
• Bake 80 °C 60 s		

- Inspect
- Soak in 40-1 (40H_2O:1NH_4OH) 10 s, rinse in DI 15 s

- Load without delay
- Evaporate:
  - 400 Å Ti
  - 200 Å Pt
  - 200 Å Pt
  - 1200 Å Au
  - 2000 Å Au
- Tape lift-off
- Remove resist in acetone
- Soak in warm acetone 5 min
- Soak in warm methanol 5 min
- Blow dry with N_2

Aluminum Etch

- Etch in Aluminum leach
- Rinse in DI 2 min
- Etch in Al leach ~50 s (Etch rate is 10 Å/s)
  Watch to see the Al clear, minimize overetch to protect small devices
- Rinse in DI 2 min
- Inspect
- Step profile (use Table below). Use arrays if large emitter mask is skipped.
- Etch in PlasmaTherm 700

Etch conditions:
Pressure 25 mTorr
Power 100 W
CF_4 rate 40 sccm
O_2 rate 3 sccm
Temperature 35 ºC
Si Etch Rate ~2 Å/s

Etch time 5 min. Target etch depth 600 Å

- Test large emitter devices
- Rinse in DI 2 min
- Blow dry in nitrogen
- Deposit 400 Å silicon nitride

Si_3N_4/FOX BONDPAD PROCESS

- Deposit FOX-15 3800 A
  - Remove FOX-15 from refrigerator, allow to warm to room temperature
  - Spin 3000 RPM 60 s
  - Hot plate bake 150 °C 1 min
  - Cure in Oxide A24-D with purging nitrogen at 300 °C for 3 hours

- Metal posts are ~4000 Å tall (check Emitter page for measured height)

- CHF_3/O_2 (trifluoromethane/oxygen) etch in PlasmaTherm 700

Etch conditions:	Pressure	40 mTorr	Power	200 W
	CHF_3 rate	50 sccm	O_2 rate	8 sccm
	Temperature	none °C		
	FOX-15 Etch Rate	6 Å/s (Pam 7/26/99)		

Etch time 5 min

- Inspect
- Step profile, left, center, and right, across 10 x 10 micron array
- Deposit 300 Å Si_3N_4 (silicon nitride) at a temperature of 250 °C to protect the FOX from AZ developer.
- Measure silicon nitride on ellipsometer, record thickness ____Å and n______

BONDPAD PROCESS (continued)

AZ1915 RESIST PROCESS – 1.5 µm resist process

Resist Coater Program 3	or	Hand Coat
• Bake 130 °C 30 s - HMDS - Bake		
• Spin 1000 rpm 3 s		• Slow Rise 130 °C
• Dispense AZ1915 400 rpm 5 s		• Spin HMDS 4000 rpm 20 s
• Spin 4000 rpm 20 s		• Spin AZ1915 4000 rpm 20 s
• Dry Spin 2500 rpm 15 s		• Hot plate prebake 120 °C 60 s
• Hot plate prebake 120 °C 60 s		

- Expose 480 ms ***reticle - TI0097, Patterns: CQW-BONDPAD6***

Resist Developer Program 2	or	Hand develop
• No post exposure bake		
• Spin 500 rpm 2 s		• No post exposure bake
• Prewet rinse 400 rpm 4.5 s		• Develop 1AZ:1H_2O 4 min
• Spray 1AZ:1H_2O 400 rpm 90 s		• Rinse DI 30 s
• DI rinse 1000 rpm 20 s		
• Spin 5000 rpm 20 s		

BOND PAD METALLIZATION

- $CF_4:O_2$ PlasmaTherm 700 etch

Etch conditions:	Pressure	40 mTorr	Power	200 W
	CH_4 rate	50 sccm	O_2 rate	8 sccm
	Temperature	none ºC	Dc bias	_____ V
	Si_3N_4 Etch Rate	250 Å/min		

Etch time 2.5 min. Target etch depth 625 Å

- Hot plate bake 80 °C 60 s
- Soak in $1HCl:1H_2O$ 30 s
- Rinse in DI 5-10 s
- Spin rinse spin dry, reduce water rinse to no more than 60 s
- Load into evaporator without delay:
- Evaporate:
  - Ti 800 Å
  - Pt 200 Å
  - Au 2000 Å
  - Au 2000 Å
- Lift-off
- Rinse in DI water, blow dry
- Inspect, final comments below

APPLIED PHYSICS LETTERS VOLUME 73, NUMBER 15 Appendix 10 12 OCTOBER 1998

Room temperature operation of epitaxially grown $\mathrm{Si/Si_{0.5}Ge_{0.5}/Si}$ resonant interband tunneling diodes

Sean L. Rommel, Thomas E. Dillon, M. W. Dashiell, H. Feng, J. Kolodzey, and Paul R. Berger[a]
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716-3130

Phillip E. Thompson and Karl D. Hobart
Naval Research Laboratory, Washington, D.C. 20375-5347

Roger Lake, Alan C. Seabaugh, Gerhard Klimeck,[b] and Daniel K. Blanks
Applied Research Laboratory, Raytheon Systems Company, Dallas, Texas 75243

(Received 1 June 1998; accepted for publication 11 August 1998)

Resonant interband tunneling diodes on silicon substrates are demonstrated using a $\mathrm{Si/Si_{0.5}Ge_{0.5}/Si}$ heterostructure grown by low temperature molecular beam epitaxy which utilized both a central intrinsic spacer and δ-doped injectors. A low substrate temperature of 370 °C was used during growth to ensure a high level of dopant incorporation. A B δ-doping spike lowered the barrier for holes to populate the quantum well at the valence band discontinuity, and an Sb δ-doping reduces the doping requirement of the *n*-type bulk Si by producing a deep n^+ well. Samples studied from the as-grown wafers showed no evidence of negative differential resistance (NDR). The effect of postgrowth rapid thermal annealing temperature was studied on tunnel diode properties. Samples which underwent heat treatment at 700 and 800 °C for 1 min, in contrast, exhibited NDR behavior. The peak-to-valley current ratio (PVCR) and peak current density of the tunnel diodes were found to depend strongly on δ-doping placement and on the annealing conditions. PVCRs ranging up to 1.54 were measured at a peak current density of 3.2 kA/cm^2. © *1998 American Institute of Physics.* [S0003-6951(98)04841-4]

Recent demonstrations of high speed and low power resonant tunneling diode/transistor circuits[1–3] have shown how the tunnel diode can boost the performance of a transistor technology. The utility of the tunnel diode has been realized since the early nineteen sixties,[4] but today tunnel diodes are used only in discrete form and for niche applications, such as high speed pulse and edge generation.[5] The drawback to the tunnel diode has long been the difficulty in controlling peak current[6] and the lack of an integrated circuit process.[4] This letter demonstrates $\mathrm{Si/Si_{0.5}Ge_{0.5}/Si}$ resonant interband tunneling diodes (RITD). The diodes studied were fabricated using molecular beam epitaxy (MBE) and incorporated a device structure suitable for integration with complementary metal-oxide-semiconductor (CMOS) or Si/SiGe heterojunction bipolar technology. With this type of device it appears that the performance benefits of an integrated Si tunnel diode/transistor technology can now be explored. The five key points to this SiGe RITD design are: (i) an intrinsic tunneling barrier, (ii) δ-doped injectors, (iii) offset of the δ-doping planes from the heterojunction interfaces, (iv) low temperature molecular beam epitaxial growth (LT-MBE), and (v) postgrowth rapid thermal annealing (RTA) for dopant activation and/or point defect reduction.

The benchmark room temperature peak-to-valley current ratio (PVCR) for alloyed Esaki tunnel diodes in Si[7] is about 4.0 and in Ge[8] is 8.3. However, the benchmark in epitaxially grown Si-based tunnel diodes is a PVCR of only 1.2 for a resonant tunneling diode which employed a relaxed buffer layer.[9] But, the highest reported PVCR of any tunnel diode (144) used an RITD configuration in the InGaAs/InAlAs material system.[10] Schematic diagrams of the RITDs examined in this study are shown in Fig. 1. The diodes studied here are similar to an RITD embodiment originally proposed by Sweeny and Xu,[11] making use of Sb and B δ-doping planes, and the Si/SiGe valence band discontinuity to achieve confined states. Figure 1(a) shows one configuration of the RITD, TD1, which employed a 4 nm undoped $\mathrm{Si_{0.5}Ge_{0.5}}$ tunneling barrier and δ-doping planes at the Si/SiGe heterointerface. Figure 1(b) shows another configuration, TD2, which was identical to TD1 described above, except that both δ-doping planes were offset from the Si/SiGe heterointerface with 1 nm of undoped Si on either side of the $\mathrm{Si_{0.5}Ge_{0.5}}$ tunneling barrier.

The variation in δ-doping placement between TD1 and TD2 was chosen because of two issues relevant to the RITDs of this study: the effects of growth interruption and dopant outdiffusion. A δ-doped layer is in essence a stop growth.

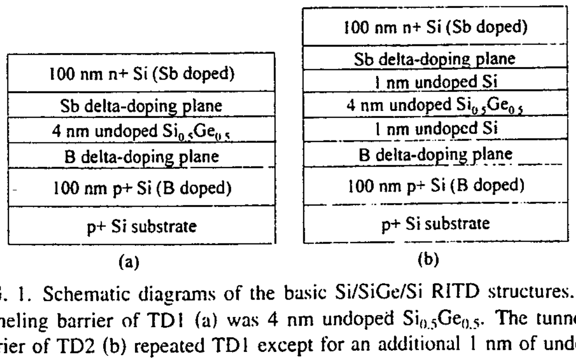


FIG. 1. Schematic diagrams of the basic Si/SiGe/Si RITD structures. The tunneling barrier of TD1 (a) was 4 nm undoped $\mathrm{Si_{0.5}Ge_{0.5}}$. The tunneling barrier of TD2 (b) repeated TD1 except for an additional 1 nm of undoped Si on either side of the $\mathrm{Si_{0.5}Ge_{0.5}}$.

---

[a] Author to whom correspondence should be addressed. Electronic mail: pberger@ee.udel.edu

[b] Present address: Jet Propulsion Laboratory, Pasadena, CA 91106-8099.

0003-6951/98/73(15)/2191/3/$15.00

© 1998 American Institute of Physics

TABLE I. Summary of the room temperature $I-V$ characteristics of $Si/Si_{0.5}Ge_{0.5}/Si$ RITDs illustrating the peak voltage, peak current density, and PVCR for TD1 and TD2, and from mesa diodes of varying diameters. Results from the best devices are presented.

Sample number	Diode Diameter (μm)	Peak Voltage (V) 700/800 °C	Peak current Density (A/cm²) 700/800 °C	PVCR 700/800 °C
TD1	18	0.35/0.12	2150/220	1.21/1.04
	50	0.55/0.25	1720/180	1.18/1.04
	75	0.68/0.36	1490/470	1.11/1.35
TD2	18	0.34/0.16	3230/520	1.54/1.18
	50	0.87/0.38	2870/430	1.52/1.26
	75	1.21/0.52	2690/470	1.48/1.30

Stop growths are commonly employed to smoothen the growth front profile and reduce heterojunction roughness.[12,13] During a stop growth, the growth rate drops considerably, but the impurity accumulation rate rises dramatically, which has been shown to quench quantum well photoluminescence.[13] Also, dopant outdiffusion from the δ-doped spike is expected to be preferentially oriented towards the undoped central $Si_{0.5}Ge_{0.5}$ spacer, rather than the highly doped outer Si injector layers, due to the concentration gradient. Furthermore, Sb diffusion[14,15] has been shown to be enhanced with increased Ge content whereas B diffusion[16] has been shown to be suppressed with the addition of Ge. Thus, the placement of the δ-doped layers offset from the $Si_{0.5}Ge_{0.5}$ spacer using undoped Si in TD2 was expected to minimize these effects and to provide a higher quality tunneling barrier with reduced defects and higher PVCR. The data presented in this study supports this supposition (see Table I).

It is well known that dopants such as Sb can segregate during MBE growth, yielding undesirably broadened δ-doping profiles. Hobart *et al.* found that dopant segregation could be suppressed by lowering the growth temperature.[17] Provided the layers are sufficiently thin, crystalline growth has been shown to occur even for extremely low temperatures (150 nm at 325 °C is a typical thickness).[18] A short postgrowth anneal has been shown to be sufficient to activate the dopants and also to anneal out point defects.[19] Therefore, in order to maximize the degeneracy of the δ-doping and bulk doping levels of the diodes in this study, a low substrate temperature of 370 °C was chosen with the understanding that an increase in point defect density might accompany the elevated doping levels.

Epitaxial growth was achieved with a specially designed MBE growth system[20] using elemental Si and Ge in *e*-beam sources, elemental Sb in a standard Knudsen cell and elemental B in a high temperature Knudsen cell. The structures were grown on 75 mm B-doped ($\rho = 0.015-0.04\ \Omega$ cm) Si(100) wafers. Prior to growth, the substrates were prepared using a cleaning technique previously described.[21] Base pressure of the MBE growth system was 5×10^{-9} Pa and typical pressure during growth was 6×10^{-7} Pa. The growths were initiated with a 1 nm undoped Si buffer layer grown at 700 °C. The substrate temperature was then lowered to 540 °C for the growth of a 70 nm B-doped p^+-Si($2\times10^{19}/cm^3$) layer. The substrate temperature was further reduced to 370 °C for the remainder of the sample growth. This included a B δ-doped layer ($7\times10^{13}/cm^2$), an undoped tunneling barrier (see Fig. 1), an Sb δ-doped layer ($3\times10^{14}/cm^2$), and a 100 nm Sb-doped n^+-Si($4\times10^{19}/cm^3$) contact layer. All regions of the sample were grown at a rate of 0.1 nm/s with the exception of the $Si_{0.5}Ge_{0.5}$ spacer, which was grown at 0.08 nm/s.

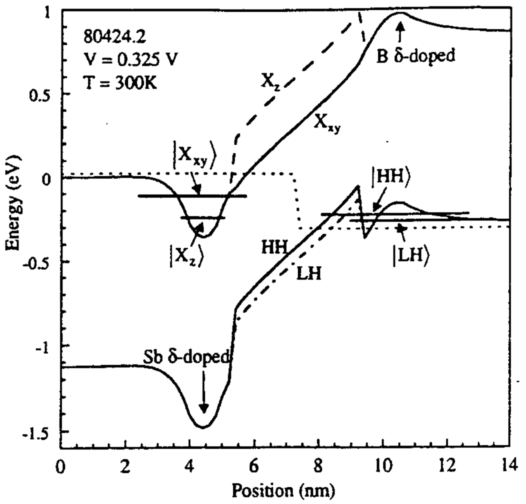


FIG. 2. Calculated band diagram of the RITD using the TD2 structure illustrated in Fig. 1(b).

Prior to device fabrication, portions of the wafers were annealed using a forming gas ambient in an AG Associates Heatpulse 610 RTA furnace at 600, 700, or 800 °C. All anneal times were held constant at 1 min. A series of Al dots with 18, 50, and 75 μm diameters were patterned on the surface of the wafers via a standard contact lithography/liftoff process. A buffered oxide etch was used prior to metallization. Using the Al dots as a self-aligned mask, a series of mesas were etched in a CF_4/O_2 plasma which resulted in some undercutting of the metal mask. Finally, an Al backside contact was thermally evaporated on all of the samples.

Figure 2 shows a calculated band diagram of tunnel diode TD2 described in Fig. 1(b). The δ-doped regions are assumed broadened over 1 nm with a dopant activation of 50%. The effective-mass Schrödinger equation and corresponding quantum charge is solved for each band and iterated to convergence with Poisson's equation. The strain in the $Si_{0.5}Ge_{0.5}$ splits the light hole and heavy hole bands by 80 meV, and it splits the X_z and X_{xy} valleys of the conduction band by 310 meV where z is the direction of crystal growth. The δ-doping provides two charge planes which determine the builtin voltage across the tunnel region. The Sb δ-doping reduces the doping requirement of the n-type bulk Si by producing a deep n^+ well. The heterojunction provides a corresponding p^+ well in the valence band, and the B δ doping reduces the barrier to holes entering the well. The light- and heavy-hole states align with the X_z electron state at 0.30 and 0.34 V, respectively. The peak current is assumed to occur in this range of voltages.

Current–voltage ($I-V$) characteristics were measured with an HP 4142 Semiconductor Parameter Analyzer and a Tektronix curve tracer. All measurements were compared on both systems for consistency. The as-grown RITDs showed no signs of NDR, and exhibited $I-V$ characteristics of leaky backward diodes.

In contrast, samples which were annealed at 700 and

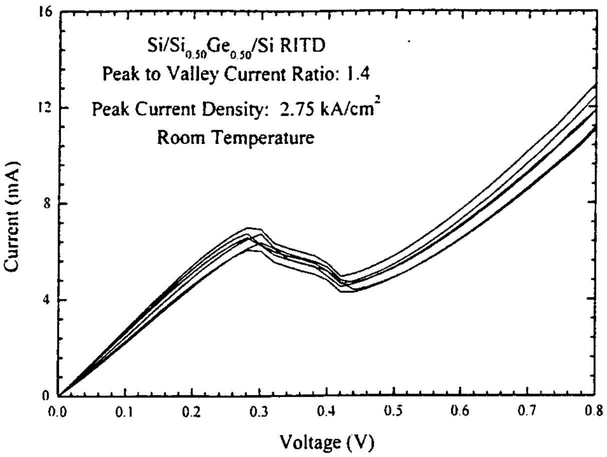


FIG. 3. $I-V$ characteristics of six representative RITDs with the TD2 structure annealed at 700 °C having 18 μm diameters which exhibit room temperature NDR (PVCR ~ 1.4 at a peak current density of 2.75 kA/cm^2).

800 °C exhibited room temperature NDR behavior (Table I and Fig. 3). TD2 was also annealed at 600 °C and only showed a plateau in its $I-V$ characteristics near the peak voltage where NDR was observed for the samples annealed at 700 and 800 °C. Furthermore, from Table I, lower PVCR and peak current density (by an order of magnitude) is observed for annealing temperatures as high as 800 vs 700 °C. A key observation is that the peak current density can be engineered during postgrowth processing with a short high temperature anneal. Optimal annealing temperatures appear to be between 600 and 800 °C, perhaps close to the 700 °C employed here.

An interesting trend was found among the samples tracking the placement of the δ-doping plane. TD1, the structure with the δ doping placed at the heterointerface, was found to have PVCRs which ranged up to 1.21 with a peak current density of 2.1 kA/cm^2 and 1.35 with a peak current density of 470 A/cm^2 when annealed at 700 and 800 °C, respectively. TD2, where the δ doping was offset 1 nm from the SiGe/Si heterointerface, exhibited PVCRs up to 1.54 with a peak current density of 3.2 kA/cm^2 and 1.30 with a peak current density of 470 A/cm^2 when annealed at 700 and 800 °C, respectively. With the exception of the 75-μm-diam diode from TD1 annealed at 800 °C, the performance of the remainder of the devices appeared to improve with the δ-doping plane offset 1 nm away from the Si/SiGe heterointerface. The aforementioned discrepancy in the data is attributed to a known radial nonuniformity in the B dopant distribution of the samples.

The peak voltages measured for all the samples shifted to higher voltage with increasing diode area due to series resistance. By plotting the peak voltage versus the peak current as a function of diode diameter, the slope of the curve yields an intrinsic series resistance of 5 Ω in the measurement setup. The extrapolated intrinsic peak voltage for TD2 annealed at 700 °C was found to be $0.3\dot{3}$ V, from the y intercept, which agrees favorably with the predicted value obtained from theoretical modeling, Fig. 2.

It should be noted that the $I-V$ characteristics for all the diodes measured in this study were repeatable and stable. Some RITDs were tested continuously for up to 1 h and without degradation in PVCR. Also, the yield of RITDs exhibiting NDR behavior is estimated over 95% from the samples annealed at 700 to 800 °C. Figure 3 shows an overlay of the $I-V$ characteristics obtained from six representative RITDs from TD2 (annealed at 700 °C) to illustrate NDR reproducibility.

In conclusion, NDR behavior at room temperature was observed in epitaxially grown Si/SiGe/Si heterostructure RITDs. The structures utilized an intrinsic $Si_{0.5}Ge_{0.5}$ tunneling barrier and δ-doped Si injectors. The placement of the δ-doping plane offset into the Si injectors may have reduced the effect of impurity diffusion of dopants into the central device region. Low temperature epitaxial growth at 370 °C allowed for a high incorporation of dopant species. Subsequent postgrowth anneals were found to activate the dopants and reduce point defect density, and an optimal anneal temperature appears to exist. NDR behavior was observed to be stable and repeatable. This study has demonstrated that SiGe RITD peak current density and PVCR can also be tailored by appropriate postgrowth processing.

The authors wish to thank D. Simons of NIST for SIMS, M. Twigg of NRL for TEM, and T. Adam, D. Aiken, C. Guedj, A. Khan, R. Jonczyk, J. Taylor, and T. Troeger of the University of Delaware; J. Cummings, B. Feyock, M. Mauk, and Z. A. Shellenbarger, of Astropower, Inc.; and S. Hui and J. Lopata of Lucent Technologies for assistance with the device fabrication and useful discussions. This work was supported by the National Science Foundation (ECS-9624160 and ECS-9622134) and by DARPA/AFOSR (F49620-96-C-0006).

[1] A. C. Seabaugh, B. Brar, T. Broekaert, G. Frazier, and P. van der Wagt, 1997 GaAs IC Symposium (unpublished), pp. 119–122.
[2] T. P. E. Broekaert, B. Brar, J. P. A. van der Wagt, A. C. Seabaugh, T. S. Moise, F. J. Morris, E. A. Beam III, and G. A. Frazier, IEEE J. Solid-State Circuits **33**, 1342 (1998).
[3] J. P. A. van der Wagt, A. C. Seabaugh, and E. A. Beam III, IEEE Electron Device Lett. **19**, 7 (1998).
[4] R. G. Swartz, 1986 IEEE International Solid-State Circuits Symposium (unpublished), pp. 278–280.
[5] Picosecond Pulse Labs, Boulder, CO.
[6] R. W. Keyes, Rev. Mod. Phys. **61**, 279 (1989).
[7] V. M. Franks, K. F. Hulme, and J. R. Morgan, Solid-State Electron. **8**, 343 (1965); N. Holonyak, Jr. (private communication).
[8] Germanium Power Devices Corporation (1996).
[9] K. Ismail, B. S. Meyerson, and P. J. Wang, Appl. Phys. Lett. **59**, 973 (1991).
[10] H. H. Tsai, Y. K. Su, H. H. Lin, R. L. Wang, and T. L. Lee, IEEE Electron Device Lett. **15**, 357 (1994).
[11] M. Sweeny and J. Xu, Appl. Phys. Lett. **54**, 546 (1989).
[12] P. R. Berger, P. K. Bhattacharya, and J. Singh, J. Appl. Phys. **61**, 2856 (1987).
[13] F-Y. Juang, P. K. Bhattacharya, and J. Singh, Appl. Phys. Lett. **48**, 290 (1986).
[14] A. N. Larsen and P. Kringhøj, Appl. Phys. Lett. **68**, 2684 (1996).
[15] P. Kringhøj, A. N. Larsen, and S. Y. Shirayev, Phys. Rev. Lett. **76**, 3372 (1996).
[16] P. Kuo, J. L. Hoyt, J. F. Gobbons, J. E. Turner, and D. Lefforge, Appl. Phys. Lett. **66**, 580 (1995).
[17] K. D. Hobart, D. J. Godbey, M. E. Twigg, M. Fatemi, P. E. Thompson, and D. S. Simons, Surf. Sci. **334**, 29 (1995).
[18] D. J. Eaglesham, H. J. Gossman, and M. Cerullo, Phys. Rev. Lett. **65**, 1227 (1995).
[19] K. D. Hobart, D. J. Godbey, and P. E. Thompson, Appl. Phys. Lett. **61**, 76 (1992).
[20] E. D. Richmond, J. G. Pellegrino, M. E. Twigg, S. Qadri, and M. T. Duffy, Thin Solid Films **192**, 287 (1990).
[21] P. E. Thompson, M. E. Twigg, D. J. Godbey, K. D. Hobart, and D. S. Simons, J. Vac. Sci. Technol. B **11**, 1077 (1993).

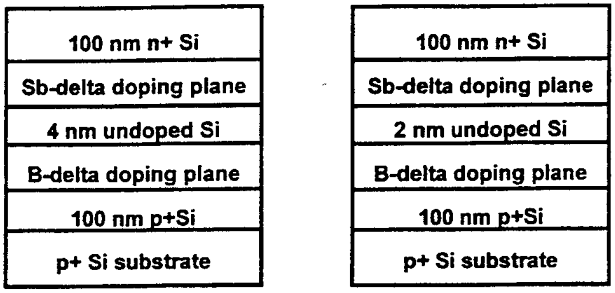
IEEE ELECTRON DEVICE LETTERS, VOL. 20, NO. 7, JULY 1999

Epitaxially Grown Si Resonant Interband Tunnel Diodes Exhibiting High Current Densities

Sean L. Rommel, Thomas E. Dillon, Paul R. Berger, Phillip E. Thompson,
Karl D. Hobart, Roger Lake, and Alan C. Seabaugh

***Abstract*— This study presents the room-temperature operation of δ-doped Si resonant interband tunneling diodes which were fabricated by low-temperature molecular beam epitaxy. Post growth rapid thermal annealing of the samples was found to improve the current–voltage ($I-V$) characteristics. Optimal performance was observed for a 600 °C 1 min anneal, yielding a peak-to-valley current ratio (PVCR) as high as 1.38 with a peak current density (J_p) as high as 1.42 kA/cm^2 for a device with a 4-nm intrinsic Si tunnel barrier. When the tunnel barrier was reduced to 2 nm, a PVCR of 1.41 with a J_p as high as 10.8 kA/cm^2 was observed. The devices withstood a series of burn-in measurements without noticeable degradation in either the J_p or PVCR. The structures presented are strain-free, and are compatible with a standard CMOS or HBT process.**

***Index Terms*—CMOS compatability, molecular beam epitaxy, negative differential resistance, rapid thermal annealing, resonant interband tunneling diodes, silicon.**



(a) (b)

Fig. 1. Schematic diagrams of the Si RITD structures. (a) The tunnel barrier of SiTD1 is 4 nm of undoped Si. (b) The tunnel barrier of SiTD2 is 2 nm of undoped Si.

I. INTRODUCTION

TUNNEL diodes are undergoing a rebirth particularly in the III–V material systems where integrated tunnel diode/transistor circuits are showing promise for mixed signal [1], [2] and memory [3], [4] applications. However, the lack of a Si-based process compatible with CMOS or SiGe heterojunction bipolar transistor (HBT) technology has impeded similar efforts in Si. Many of the earliest Si-based tunnel diodes, including the Si Esaki diode with the highest reported peak-to-valley current ratio (PVCR) of 4.0 [5], were fabricated via an alloying process. While this technique was adept at fabricating discrete devices, it was not suited for integrated circuit processing [6]. Valient attempts were made in the early 1990's to fabricate Si/SiGe resonant tunneling diodes (RTD's), but the lack of a material capable of achieving a conduction band offset prevented the realization of RTD's with PVCR's much greater than 1.2 [7]. A recent letter by the authors presented the room-temperature operation of δ-doped Si/Si$_{0.5}$Ge$_{0.5}$/Si resonant interband tunneling diodes (RITD's) which were fabricated epitaxially by low-temperature molecular beam epitaxy (LT-MBE) [8]. As-grown samples showed no evidence of negative differential resistance (NDR). However, heat treatments in a rapid thermal annealing (RTA) furnace used for dopant activation and point defect reduction resulted in NDR behavior. PVCR's as high as 1.5 at a peak current density of 3.2 kA/cm^2 were reported for a 1-min anneal at 700 °C. Beyond this temperature, the current density was found to abruptly decrease by an order of magnitude. This letter builds upon the previous study, presenting epitaxially fabricated δ-doped RITD structures consisting purely of Si which exhibit NDR at room temperature.

The key differences between the RITD's studied here and the previously reported Si/SiGe/Si RITD's are 1) a purely Si tunnel barrier eliminates critical thickness issues and strain associated with SiGe alloys, 2) strain relaxation during the RTA heat treatments is avoided, and 3) quantum confinement is not aided by the presence of a heterojunction discontinuity. The two structures presented, which will be referred to as SiTD1 and SiTD2, are shown in Fig. 1. The only variation between the two structures is the tunnel barrier thickness, which is 4 nm of undoped Si for SiTD1 and 2 nm of undoped Si for SiTD2. As with the previously reported Si/Si$_{0.5}$Ge$_{0.5}$/Si RITD's, B and Sb δ-doped injectors are used to create confined states in the valence and conduction bands, respectively. Fig. 2 shows a calculated band diagram for SiTD1. The δ-doped regions are assumed broadened over 1 nm with a dopant activation of 50%. The diagram was generated by solving the effective-mass Schrödinger equation and corresponding quantum charge for each band and iterating to convergence with Poisson's equation. As evidenced in this diagram, the built-in voltage drops across the intrinisic region. This region,

---

Manuscript received December 18, 1998; revised April 5, 1999. This work was supported by DARPA/AFOSR (F49620-96-C-0006) and by the National Science Foundation (ECS-9624160 and ECS-9622134).

S. L. Rommel, T. E. Dillon, and P. R. Berger are with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA.

P. E. Thompson and K. D. Hobart are with the Naval Research Laboratory, Washington, DC 20375 USA.

R. Lake and A. C. Seabaugh are with the Applied Research Laboratory, Raytheon Systems Company, Dallas, TX 75243 USA.

Publisher Item Identifier S 0741-3106(99)05665-7.

0741-3106/99$10.00 © 1999 IEEE

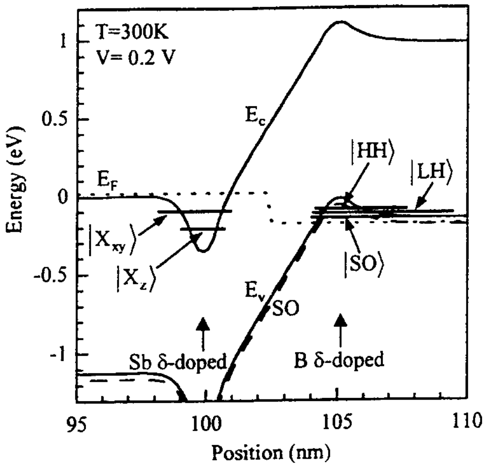


Fig. 2. Calculated energy band diagram and resonant states of SiTD1. The Sb and B activation of the δ-doped regions is assumed to be 50%. X_z denotes the conduction band minimums along the k_z axis of the Brillouin zone, and X_{xy} denotes the conduction band minimums along the k_x and k_y axes where z is the growth direction. SO denotes the split-off valence band-edge.

TABLE I

SUMMARY OF THE ROOM TEMPERATURE $I-V$ CHARACTERISTICS OF THE Si RITD'S FOR SiTD1 AND SiTD2. RESULTS FROM 18 μm DIAMETERS DIODES FROM SUBSTRATES ANNEALED AT 1 min ARE SHOWN. A PVCR ENTRY OF 1.00 INDICATES THAT THE DEVICE EXHIBITED A PLATEAU IN THE FORWARD DIRECTION RATHER THAN NDR BEHAVIOR

	T (°C)	500	550	600	650	700
SiTD1	J_p (kA/cm^2)	1.60	0.61	1.36	0.87	0.17
	J_v (kA/cm^2)	——	0.52	0.98	0.65	0.15
	PVCR	1.00	1.15	1.38	1.35	1.15
SiTD2	J_p (kA/cm^2)	32.00	15.8	9.40	1.40	0.46
	J_v (kA/cm^2)	27.0	14.8	6.28	1.28	0.41
	PVCR	1.17	1.07	1.45	1.1	1.1

therefore, is also treated as the tunnel barrier. As Si has an indirect bandgap, the peak current in this structure is attributed to phonon assisted tunneling between the X_{xy} electron and light hole states [9]. The NDR in this structure results from a decrease in the tunneling probability with applied bias as these bands uncross.

II. EXPERIMENTAL SETUP

Epitaxial growth was achieved with a specially designed MBE growth system [10] using elemental Si and Ge in e-beam sources, elemental Sb in a standard Knudsen cell and elemental B in a high-temperature Knudsen cell. The structures were grown on 75-mm B-doped (ρ = 0.015–0.04 Ω·cm) Si(100) wafers. Prior to growth, the substrates were prepared using a cleaning technique previously described [11]. The growths were initiated with a 2-nm undoped Si buffer layer grown at 700 °C. The substrate temperature was then lowered to 540 °C for the growth of a 70-nm B-doped p$^+$-Si (2×10^{19}/cm^3) layer. The substrate temperature was further reduced to 370 °C for the remainder of the sample growth. This included a B δ-doped layer (7×10^{13}/cm^2), an undoped tunneling barrier (see Fig. 1), an Sb δ-doped layer (3×10^{14}/cm^2), and a 100-nm Sb-doped n$^+$-Si (4×10^{19}/cm^3) contact layer. All regions of the sample were grown at a rate of 0.1 nm/s. Furthermore, with the exception of the lack of Ge in SiTD1 and SiTD2, their growth and doping was essentially identical to that of the previously reported Si/Si$_{0.5}$Ge$_{0.5}$/Si RITD's.

A series of samples were annealed under a 15% H$_2$ 85% N$_2$ ambient in an AG Associates Heatpulse 610 Rapid Thermal Annealing Furnace. The anneal time of all samples was fixed at one minute. Anneal temperatures of 500, 550, 600, 650, and 700 °C were employed in this study. Control samples without heat treatment were also fabricated for SiTD1 and SiTD2.

Al dots with diameters of 5, 10, 18, 50, and 75 μm were patterned on the surface of the wafer via a standard contact lithography/liftoff process. The samples were dipped in a buffered oxide etch prior to metallization to remove the native oxide. The Al dots served as a mask when the samples were etched in a CF$_4$/O$_2$ plasma for mesa isolation. Finally, an Al backside contact was thermally evaporated to complete the device fabrication.

III. RESULTS AND DISCUSSION

Room temperature current–voltage ($I-V$) chacteristics were measured both with an HP 4142 Semiconductor Parameter Analyzer and a Textronix curve tracer. In contrast to the previous study, NDR behavior was observed in the control samples which were not heat treated. However, NDR was only evident in the 5- and 10-μm diameter diodes of the control samples, and the observed PVCR from these samples was barely greater than one. Post-growth heat treatments substantially improved the device performance.

Table I summarizes the peak current density (J_p), valley current density (J_v), and PVCR of 18 μm diameter diodes resulting from each anneal temperature employed on SiTD1 and SiTD2. The largest combination of J_p and PVCR occurred after a 600 °C, 1-min anneal for both SiTD1 and SiTD2; a PVCR of 1.38 with a J_p of 1.36 kA/cm^2, and a PVCR of 1.45 with a J_p of 9.4 kA/cm^2 were observed for SiTD1 and SiTD2, respectively. Anneal temperatures above this optimal value led to lower values of J_p. It should also be noted that the degradation of Si RITD's occurred at an anneal temperature 100 °C below that of the Si/Si$_{0.5}$Ge$_{0.5}$/Si RITD study [8], possibly because the diffusion of B is more rapid in bulk Si than in Si$_{1-x}$Ge$_x$ alloys [12].

Spacer thickness had a significant influence on J_p and v. Fig. 3 shows an overlay of the $I-V$ characteristics of 10 μm diameter diodes from SiTD1 and SiTD2, both annealed at a temperature of 600 °C for 1 min. The current density of SiTD2 (10.8 kA/cm^2) is almost an order of magnitude larger than that of SiTD1 (1.42 kA/cm^2). Since the devices were grown under identical conditions with the exception of spacer thickness, the elevated current density of SiTD2 must be due to its thinner spacer. The fact that all entries from SiTD1 in Table I have smaller J_p and J_v values than corresponding entries from SiTD2 reinforces this conclusion. However, a reduction in spacer thickness did not yield as sharp of an increase in current density as the authors originally suspected. The may suggest that the depletion region for these structures extends beyond the δ-doping planes. With proper adjustments to the doping levels and to growth parameters, this problem may be avoided.

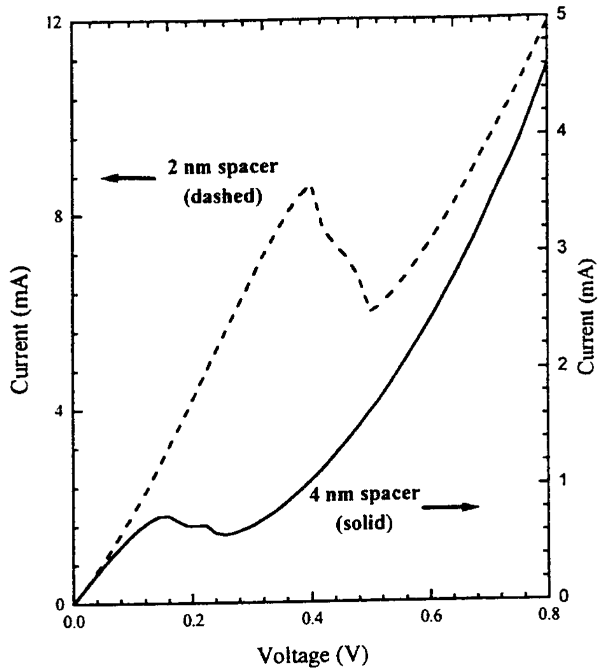


Fig. 3. I–V characteristics of Si RITD's with 4-nm (SiTD1) and 2-nm (SiTD2) tunnel barriers. The current from the 2-nm structure and 4-nm structures are plotted on the left and right axes, respectively. The diodes shown have a 10-μm diameter and were annealed at 600 °C for 1 min. The 4-nm device had a PVCR of 1.42 with a J_p of 1.42 kA/cm^2. The 2-nm device had a PVCR of 1.41 with a J_p of 10.8 kA/cm^2.

The variation in peak voltage evident in Fig. 3 is attributed to the series resistance of the Al contacts.

A burn-in study was performed on a portion of SiTD2 annealed at 600 °C. The devices on this sample were cycled between -1.5 and 1.5 V at a rate of 18 cycles/s for the first 20 500 cycles. The rate was then increased to 42 cycles/s for the duration of the measurements. A 0.47% reduction in J_p and J_v was observed over the first 20 500 cycles. The $I-V$ charactertistics then remained essentially unchanged after 83 800 cycles.

IV. CONCLUSION

In conclusion, Si RITD's were fabricated epitaxially and found to exhibit NDR behavior at room temperature. PVCR's up to 1.42 with a J_p as high as 1.42 kA/cm^2 and up to 1.41 with a J_p as high as 10.8 kA/cm^2 were observed after a 1-min 600 °C anneal for 4- and 2-nm undoped Si spacers, respectively. The current density of the RITD's was found to depend on the thickness of the undoped Si spacer, allowing for engineering of the peak current density for specific device applications. Optimal anneal conditions may be affected by the interdiffusion of dopants, as device degradation occurred 100 °C below the 700 °C, 1-min optimal anneal previously reported for Si/SiGe/Si RITD's. The devices withstood a burn-in study which demonstrated their reliability for circuit operation. The device structure, as well its low temperature fabrication process makes it a strong candidate for integration in a CMOS or HBT fabrication line.

ACKNOWLEDGMENT

The authors would like to thank C. Guedj (University of Delaware) for X-ray diffraction, D. S. Simons (NIST) for secondary ion mass spectrometry, M. Twigg (NRL) for high-resolution transmission electron microscopy, and Astropower Inc. for help in device fabrication. NEMO software was used to calculate the structure of the band diagrams. In addition, the authors wish to recognize J. Kolodzey (University of Delaware) and G. Pomerenke (DARPA/ETO) for their guidance and support.

REFERENCES

[1] T. P. E. Broekaert, B. Brar, J. P. A. van der Wagt, A. C. Seabaugh, T. S. Moise, F. J. Morris, E. A. Beam, III, and G. A. Frazier, "A monolithic 4-bit 2-Gsps resonant tunneling analog-to-digital converter," *IEEE J. Solid State Circuits*, vol. 33, pp. 1342–1349, 1998.

[2] A. C. Seabaugh, B. Brar, T. Broekaert, F. Morris, and G. Frazier, "Resonant tunneling analog/mixed signal circuit technology," *Top. Workshop Heterostructure Microelectronics for Information Systems Applications*, 1998, pp. 14–15.

[3] J. P. A. van der Wagt, A. C. Seabaugh, and E. A. Beam, III, "RTD/HFET low standby power SRAM gain cell," *IEEE Electron Device Lett.*, vol. 19, pp. 7–9, 1998.

[4] J. P. A. van der Wagt, "Tunneling-based SRAM," *Proc. IEEE*, vol. 87, pp. 571–595, 1999.

[5] V. M. Franks, K. F. Hulme, and J. R. Morgan, "An alloy process for making high current density Si tunnel diode junctions," *Solid State Electron.*, vol. 8, pp. 343–344, 1965.

[6] R. G. Swartz, "In perspective: The tunnel diode," in *1986 IEEE Int. Solid-State Circuits Symp.*, pp. 278–280.

[7] K. Ismail, B. S. Meyerson, and P. J. Wang, "Electron resonant tunneling in Si/SiGe double barrier diodes," *Appl. Phys. Lett.*, vol. 59, pp. 973–975, 1991.

[8] S. L. Rommel, T. E. Dillon, III, M. W. Dashiell, H. Feng, J. Kolodzey, P. R. Berger, P. E. Thompson, K. D. Hobart, R. Lake, A. C. Seabaugh, G. Klimeck, and D. K. Blanks, "Room temperature operation of epitaxially grown Si/Si$_{0.5}$Ge$_{0.5}$/Si resonant interband tunneling diodes," *Appl. Phys. Lett.*, vol. 73, pp. 2191–2193, 1998.

[9] N. Holonyak, I. A. Lesk, R. N. Hall, J. J. Tiemann, and H. Ehrenreich, "Direct observation of phonons during tunneling in narrow junction diodes," *Phys. Rev. Lett.*, vol. 3, pp. 167–168, 1959.

[10] E. D. Richmond, J. G. Pellegrino, M. E. Twigg, S. Qadri, and M. T. Duffy, "Unique properties of molecular beam epitaxy on sapphire using *in-situ* high temperature substrate annealing compared with chemically vapor deposited silicon on sapphire," *Thin Solid Films*, vol. 192, pp. 287–294, 1990.

[11] P. E. Thompson, M. E. Twigg, D. J. Godbey, K. D. Hobart, and D. S. Simons, "Low-temperature cleaning processes for Si molecular beam epitaxy," *J. Vac. Sci. Technol. B*, vol. 11, pp. 1077–1082, 1993.

[12] P. Kuo, J. L. Hoyt, J. F. Gibbons, J. E. Turner, and D. Lefforge, "Effects of strain on boron diffusion in Si and Si$_{1-x}$Ge$_x$," *Appl. Phys. Lett.*, vol. 66, pp. 580–582, 1995.

Appendix 12

Si-Based Interband Tunneling Devices For High-Speed Logic and Low Power Memory Applications

Sean L. Rommel, Thomas E. Dillon, Paul R. Berger, Roger Lake,[a] Phillip E. Thompson,[b] Karl D. Hobart,[b] Alan C. Seabaugh,[a]
and David S. Simons[c]
University of Delaware, Department of Electrical and Computer Engineering, Newark, DE 19716-3130
[a]Raytheon Systems, Dallas, TX 75243, [b]Naval Research Laboratory, Washington, D.C. 20375,
[c]NIST, Gaithersburg, MD 20899

Tunnel diode/transistor logic (TDTL), realized to date in III-V material systems only, enhances any transistor technology [1]. Memory circuits incorporating low current density double barrier resonant tunneling diodes (DBRTDs) [2] and resonant interband tunneling diodes (RITDs) [3] have been demonstrated with fewer transistors and lower power dissipation than conventional CMOS SRAM or DRAM circuitry. High current density DBRTDs, in contrast, have been shown to successfully improve the speed and power of logic circuitry such as multiplexers [4] and analog-to-digital converters [5]. However, such technology will never reach mainstream CMOS or SiGe HBT technology unless a suitable Si-based negative differential resistance (NDR) device can be developed.

This study extends the preliminary work of the authors [6], presenting improved epitaxially grown $Si/Si_{0.5}Ge_{0.5}/Si$ resonant interband tunnel diodes (RITDs) with current densities which exceed any previously reported [7] for a Si-based NDR device. For the first time, the needs of Si-based TDTL circuits are beginning to be addressed. Two new classes of Si-based NDR devices are also reported here: *Si-only* RITDs and $Si/Si_{0.5}Ge_{0.5}$ heterojunction Esaki tunnel diodes with a digitally graded superlattice (DG-SL). The key to achieving room temperature NDR is developing tight control of dopant profiles during epitaxial growth. The structures presented here make use of δ-doping planes of B and Sb to reduce the doping requirements of the bulk regions and also to define the tunnel barrier. The use of a low substrate temperature (370°C) during molecular beam epitaxial (MBE) growth leads to higher dopant incorporation and minimizes the segregation of dopants, such as Sb, during growth [8]. Post growth rapid thermal anneals (RTA) were employed to activate dopants and reduce point defect density. An intrinsic tunnel barrier was incorporated in all designs to reduce gap states and band tails in order to reduce the excess valley current. As will be shown in this study, precise control of the dopant profile, the material content of the tunnel barrier, and the device geometry are all critical to engineering the desired current density.

Fig. 1 presents a schematic diagram of a design for a high current density RITD. The intrinsic tunnel region consists of a 2 nm layer of strained, pseudomorphic $Si_{0.5}Ge_{0.5}$ cladded on either side with 1 nm of Si to separate the δ-doped layers from the heterointerface. The narrow bandgap $Si_{0.5}Ge_{0.5}$ is used to enhance the current density. The depletion region and tunnel barrier are confined between the two δ-doped planes. Fig. 2 shows the calculated band diagram and resonant states of the device assuming 1 nm diffusive broadening of the δ-doped regions. The quantum charge was calculated for each band from the effective mass Schrödinger equation and iterated to convergence with Poisson's equation. Secondary ion mass spectrometry (SIMS) depth profiling was performed on this structure after a 700°C 1 minute RTA which indicated significant segregation of the Sb during growth. A second band diagram of this structure was generated using the SIMS data, and is shown in Fig. 3. The later band diagram suggests that the quantum confinement of the Sb δ-doping spike is lost due to dopant segregation during growth. As SIMS depth resolution is known to suffer from knock-on effects, possibly exaggerating the degree of dopant segregation, an interpolation between the two band diagrams of Figs. 2 and 3 would be reasonable. If in fact there is quantum confinement due to the Sb δ-doping, the well is certainly more shallow than indicated by the ideal band diagram of Fig. 2. Fig. 4 shows a curve tracer plot of the I-V characteristics from a representative $Si/Si_{0.5}Ge_{0.5}/Si$ RITD which exhibited a peak-to-valley current ratio (PVCR) up to 2.05 with a peak current density of 22 kA/cm^2. To the knowledge of the authors, this current density exceeds any previously reported for a Si-based NDR structure.

A *Si-only* version of the above device resulted in an order of magnitude decrease in the current density which is still large for a *Si-only* tunnel diode. A schematic of the *Si-only* device is shown in Fig. 5 and a calculated band diagram using the expected dopant profile is shown in Fig. 6. Optimal device performance was found to occur for a 625°C 1 minute RTA for the *Si-only* RITDs. Fig. 7 shows representative I-V characteristics of these RITDs captured by a Hewlett-Packard 4142 Semiconductor Parameter Analyzer. A PVCR up to 1.39 with a current density of 3.1 kA/cm^2 was observed.

Due to the Sb segregation, the depletion region and tunnel barrier of both the $Si/Si_{0.5}Ge_{0.5}/Si$ and the *Si-only* RITD widen several nm beyond that of the design. The authors have reason to believe that greater control over the Sb segregation can be obtained by a lower growth temperature. Furthermore, simulations show that the δ-doped planes can be placed closer and still maintain the built in voltage between them. Since tunnel current depends exponentially on barrier width, there is room for one to two orders of magnitude increase in the current density for both the $Si/Si_{0.5}Ge_{0.5}/Si$ and the *Si-only* RITD.

A final structure, a $Si/Si_{0.5}Ge_{0.5}$ heterojunction Esaki diode with a DG-SL, is shown in Fig. 8. The structure was designed to take advantage of the natural band offsets between Si and $Si_{0.5}Ge_{0.5}$, which lowers the doping requirements on the P^+ side of the junction. A 10 nm DG-SL was incorporated on the P^+ side of the junction to reduce the barrier for holes to populate the valence band well. In order to *lower* the current density, a 4 nm intrinsic Si spacer was placed between the P^{++} and N^{++} layers. The band diagram of this structure is shown in Fig. 9. Fig. 10 shows the I-V characteristics of representative tunnel diodes which had a PVCR up to 1.2 with a peak current density of 7.5 A/cm^2, almost 4 orders of magnitude smaller than the $Si/Si_{0.5}Ge_{0.5}/Si$ RITDs. The low current density of this structure makes it a potential design to optimize for memory applications.

In conclusion, the authors have demonstrated a family of epitaxially fabricated Si-based NDR devices operating at room temperature. The precise control of dopant profiles as a result of the low temperature growth and post growth heat treatments enables the engineering of current densities, something previously unattainable for Si-based NDR structures. The authors have further demonstrated devices which have current densities exceeding any previously reported in Si-based NDR structures. With further adjustments to the growth conditions and device geometries, this value is expected to exceed 10^5 A/cm^2. Low current density devices suitable for memory applications have also been demonstrated. With this new fabrication technology, Si-based TDTL circuitry may become a reality.

0-7803-4777-3/98/$10.00 (c) 1998 IEEE

REFERENCES

[1] A. Seabaugh, B. Brar, T. Broekaert, G. Frazier, F. Morris, P. van der Wagt, and E. Beam, "Resonant tunneling circuit technology: Has it arrived?," *1997 GaAs IC Symposium*, 1997.

[2] J. P. van der Wagt, A. Seabaugh, and E. Beam, "RTD/HFET low standby power SRAM gain cell," *1996 IEDM Technical Digest*, pp. 425-428.

[3] J. Shen, G. Kramer and R. Tsui, "Static random access memories based on resonant interband tunneling diodes in the InAs/GaSb/AlSb material system," *IEEE Elec. Dev. Lett.*, vol. 16, pp. 178–180, 1996.

[4] H. L. Chan, S. Mohan, P. Mazumder, G. I. Haddad, "Compact multiple-valued multiplexers using negative differential resistance devices," *IEEE J. of Solid-State Circuits*, vol. 31, pp. 1151–1156, 1996.

[5] T. Broekaert, B. Brar, J. P. van der Wagt, A. Seabaugh, F. Morris, T. Moise, E. Beam, and G. Frazier, "A monolithic 4-bit 2-gsps resonant tunneling analog-to-digital converter," *IEEE J. Solid-State Circ.*, vol. 33, pp. 1342–1349, 1998.

[6] S. L. Rommel, T. E. Dillon, M. W. Dashiell, H. Feng, J. Kolodzey, P. R. Berger, P. E. Thompson, K. D. Hobart, R. Lake, A. C. Seabaugh, G. Klimeck, and D. Blanks, "Room temperature operation of epitaxially grown Si/SiGe/Si resonant interband tunneling diodes," *to appear in Appl. Phys. Lett.*, vol. 73, 1998.

[7] V. M. Franks,K. F. Hulme, and J. R. Morgan, "An alloy process for making high current density Si tunnel diode junctions," *Solid State Electronics*, vol. 8, pp. 343-344, 1965.

[8] K. D. Hobart, D. J. Godbey, M. E. Twigg, M. Fatemi, P. E. Thompson, D. S. Simons, *Surf. Sci.*, vol. 334, pp. 29-35, 1995.

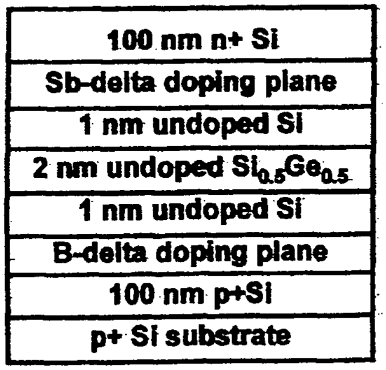


Fig. 1. Schematic diagram of the $Si/Si_{0.5}Ge_{0.5}/Si$ RITD structure.

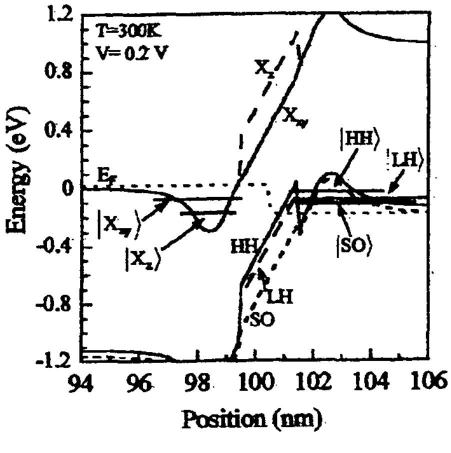


Fig. 2. Calculated band diagram and resonant states of the high current density $Si/Si_{0.5}Ge_{0.5}/Si$ RITD. 1 nm diffusive broadening of the δ-doped profiles were assumed for this calculation.

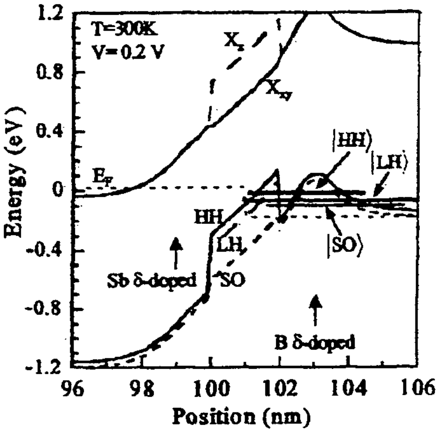


Fig. 3. Calculated band diagram of the high current density $Si/Si_{0.5}Ge_{0.5}/Si$ RITD using the SIMS Sb profile. Note that confinement on the Sb side of the junction has lost confinement because of segregation during growth.

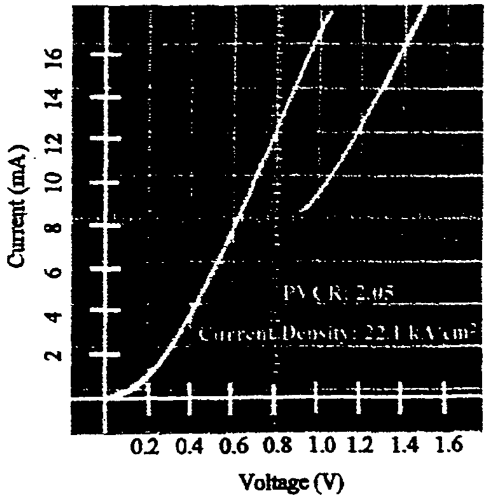


Fig. 4. I-V characteristics of the high current density $Si/Si_{0.5}Ge_{0.5}/Si$ RITD structure annealed at 700°C having an 18 μm diameter which exhibits room temperature NDR (PVCR ~ 2.05 at a peak current density of 22 kA/cm²).

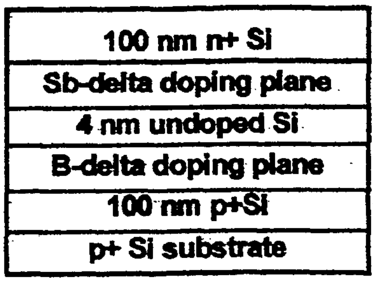


Fig. 5. Schematic diagram of the *Si-only* RITD structure.

0-7803-4777-3/98/$10.00 (c) 1998 IEEE

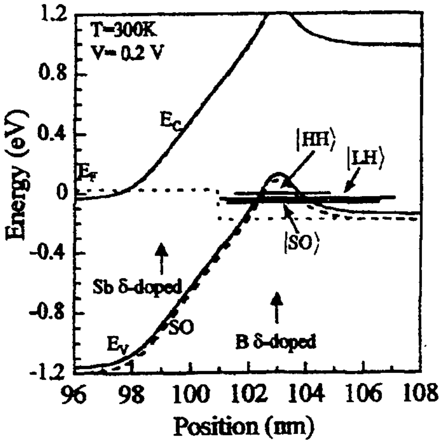


Fig. 6. Calculated band diagram and resonant states of the *Si-only* RITD of Fig. 5 using the expected Sb doping profile.

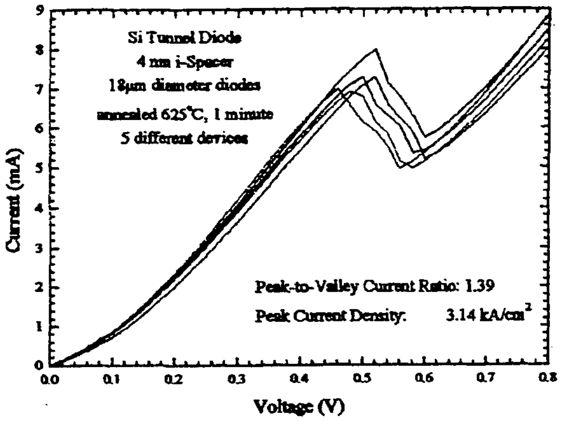


Fig. 7. I-V characteristics of six representative *Si-only* RITDs annealed at 625°C having 18 μm diameters which exhibit room temperature NDR (PVCR ~ 1.4 at a peak current density of 3.14 kA/cm^2).

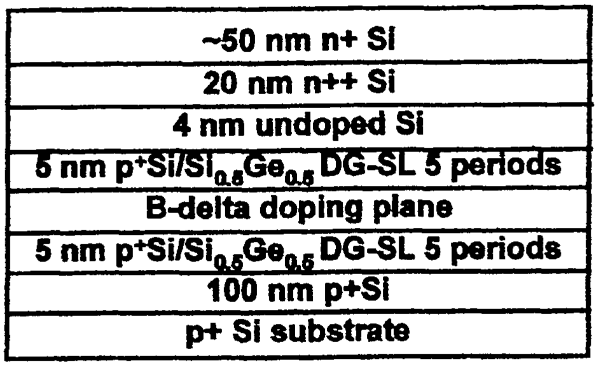

~50 nm n+ Si
20 nm n++ Si
4 nm undoped Si
5 nm p⁺Si/Si0.5Ge0.5 DG-SL 5 periods
B-delta doping plane
5 nm p⁺Si/Si0.5Ge0.5 DG-SL 5 periods
100 nm p+Si
p+ Si substrate

Fig. 8. Schematic diagram of the Si/Si$_{0.5}$Ge$_{0.5}$ heterojunction Esaki diode with a DG-SL structure.

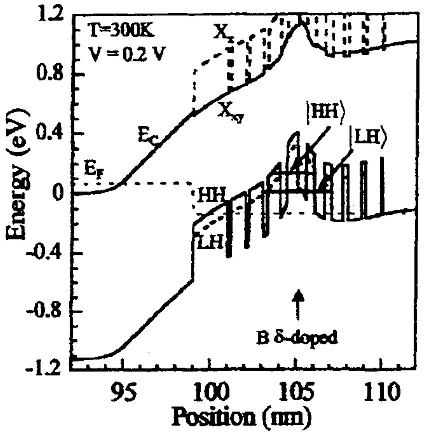


Fig. 9. Calculated band diagram and resonant states of the low current density DG-SL heterojunction Esaki diode.

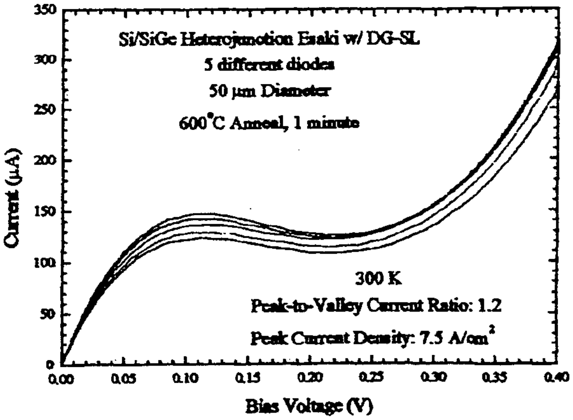


Fig. 10. I-V characteristics of five representative DG-SL heterojunction Esaki diodes annealed at 600°C having 50 μm diameters which exhibit room temperature NDR (PVCR ~ 1.2 at a peak current density of 7.5 A/cm^2).

0-7803-4777-3/98/$10.00 (c) 1998 IEEE

APPLIED PHYSICS LETTERS VOLUME 75, NUMBER 9 30 AUGUST 1999

Si resonant interband tunnel diodes grown by low-temperature molecular-beam epitaxy

Phillip E. Thompson,[a] Karl D. Hobart, Mark E. Twigg, and Glenn G. Jernigan
Naval Research Laboratory, Washington, DC 20375-5347

Thomas E. Dillon, Sean L. Rommel, and Paul R. Berger
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716-3130

David S. Simons and Peter H. Chi
National Institute of Standards and Technology, Gaithersburg, Maryland 20899

Roger Lake and Alan C. Seabaugh
Applied Research Laboratory, Raytheon Systems Company, Dallas, Texas 75243

(Received 20 April 1999; accepted for publication 2 July 1999)

Si resonant interband tunnel diodes that demonstrate negative differential resistance at room temperature, with peak-to-valley current ratios greater than 2, are presented. The structures were grown using low-temperature (320 °C) molecular-beam epitaxy followed by a postgrowth anneal. After a 650 °C, 1 min rapid thermal anneal, the average peak-to-valley current ratio was 2.05 for a set of seven adjacent diodes. The atomic distribution profiles of the as-grown and annealed structures were obtained by secondary ion mass spectrometry. Based on these measurements, the band structure was modeled and current–voltage trends were predicted. These diodes are compatible with transistor integration. © *1999 American Institute of Physics.*
[S0003-6951(99)01935-X]

There is current interest in tunnel diodes, which have characteristics very useful to the circuit designer for applications such as embedded memory and signal processing. Recent review articles[1–3] discuss potential applications of this family of devices in Si. The first tunnel diode, the Esaki diode,[4] which is an intrinsically simple device consisting of a degenerately doped p/n junction, has not found wide use in integrated circuits because of the lack of an epitaxial formation process. Epitaxial processes have been used for Si-based resonant tunnel diodes (RTD), which have demonstrated hole tunneling,[5–10] where the layers are grown on a Si substrate such that the Si barriers are relaxed and the SiGe well layer is compressively strained, and electron tunneling,[11,12] where the barrier layers are relaxed SiGe and the Si well layer is under tensile stress. None of the hole RTDs have reported negative differential resistance (NDR) above a temperature of 77 K, which severely restricts the range of their application. The electron RTD reported by Ismail[11] had a peak-to-valley current ratio (PVCR) of 1.2 at 300 K, while the device reported by Matutinovic-Krstelj[12] only had NDR at temperatures less than 220 K. Additionally, the electron RTDs are not compatible with complimentary metal oxide silicon (CMOS) or heterojunction bipolar transistor (HBT) integration since they require a thick (>1 μm) relaxed SiGe buffer layer. A recent device reported by Sardela[13] employing B δ-doped layers in Si had NDR at room temperature, with a PVCR of 1.1 and a peak current density of 4.8 $\times 10^3$ A/cm^2. The exciting feature of the B δ-doped device is that the structure is compatible with CMOS and HBT integration, although the PVCR must be increased for the device to be useful. The search, then, is for a Si-based tunnel diode which has improved electrical characteristics and which is compatible with transistor integration.

In our initial study[14,15] we investigated SiGe resonant interband tunnel diodes (RITD), similar to RITD embodiment originally proposed by Sweeney and Xu.[16] By using low-temperature (370 °C) molecular-beam epitaxy (MBE), we were able to fabricate Si/$Si_{0.5}Ge_{0.5}$ RITDs having, at room temperature, a PVCR of 2.05. By replacing the $Si_{0.5}Ge_{0.5}$ with Si, we were able to produce a Si tunnel diode having a PVCR of 1.41.[15,17] The performance of the tunnel diodes was found to depend strongly on the δ-doping spacing and on the postgrowth annealing conditions. In this letter, we will demonstrate Si RITDs which have PVCR values greater than 2 without the added complexity of Ge.

The Si RITDs were grown on a 75 mm Si (100) B-doped (0.015–0.04 Ω cm) substrate using solid-source MBE. Prior to entry into the specially designed MBE growth system,[18] the Si substrate was cleaned using a procedure previously described,[19] which resulted in a stable, hydrogen-terminated surface. Si was deposited by e-beam evaporation. The dopants, B and Sb, were obtained by evaporation of elemental sources in Knudsen cells. The substrate temperature during growth was monitored by an optical pyrometer which was calibrated by observing the eutectic temperatures of Au/Si (363 °C) and Al/Si (577 °C) on equivalent substrates.

The device structure employed in this study is shown in Fig. 1. The key features of the growth which differentiate this from earlier structures are: (1) lower substrate temperature (320 °C) during growth, which was employed to minimize the segregation of the dopants; (2) increased doping in the $p^+\delta$ layer; and (3) undoped layers surrounding the

---

[a] Corresponding author. Electronic mail: thompson@estd.nrl.navy.mil

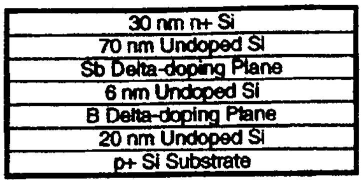


FIG. 1. Schematic diagram of the nominal Si RITD structure.

δ-doped planes to enhance carrier confinement. Growth was initiated at 650 °C to ensure removal of the hydrogen surface termination. After the growth of 10 nm of undoped Si, the substrate temperature was reduced to 450 °C while an additional 10 nm layer of undoped Si was grown. Si growth was interrupted and, concurrent with the deposition of B ($3 \times 10^{14}/\mathrm{cm}^2$) in a δ-doped layer, the temperature of the substrate was further reduced to 320 °C, which was the final temperature of the substrate during the remainder of the growth. A 6 nm undoped Si spacer layer was grown followed by a second growth interrupt for the deposition of an Sb δ-doped layer ($3 \times 10^{14}/\mathrm{cm}^2$). The final layers were a spacer layer of 70 nm of undoped Si followed by 30 nm of Sb-doped ($>10^{20}/\mathrm{cm}^3$) Si for the top Ohmic contact. The Si growth rate was 0.1 nm/s, with the exception of the δ-doped regions during which the Si flux was shuttered. Prior to device fabrication, portions of the samples were annealed to determine if postgrowth anneal improved the device characteristics, analogous to our previous RITD work.[14–16] Rapid thermal anneal (RTA) was employed using temperatures of 600, 650, or 700 °C for a period of 1 min.

Atomic profiles were obtained by secondary ion mass spectrometry (SIMS) using a high-performance magnetic sector secondary ion mass spectrometer. The net impact energy of the primary beam, 3 keV O_2^+, was selected in order to minimize profile broadening by ion-beam mixing. Depth scales were obtained from stylus profilometry (±3% uncertainty). The atomic carrier concentrations of B and Sb were calibrated with implant standards (±10% uncertainty). The electrically active portion of the RITD is the region which includes the n^+ layer, the p^+ layer, and the undoped spacer between them. This region is presented in Fig. 2, for the as-grown sample (solid lines) and the sample which underwent RTA for 1 min at 700 °C (dashed lines). Some of the width observed in the doping profiles is due to ion-beam mixing during the SIMS measurement. The portion of the dopant profiles extending into the sample are more affected by the SIMS measurement. It is observed that the B δ-doped profile was affected by the RTA. The half width at half maximum, toward the surface, increased from 3 to 5 nm. There was a corresponding reduction in the peak B concentration from 3.6×10^{20} to $2.5 \times 10^{20}/\mathrm{cm}^3$. The differences observed between the as-grown and annealed Sb δ-doped profiles are within the mutual uncertainties of the SIMS measurements. Based on the atomic distributions obtained by SIMS, the band diagrams for the RITD, as-grown, and after RTA for 1 min at 700 °C, were generated by solving the effective-mass Schrödinger equation and corresponding quantum charge for each band and iterating to convergence with Poisson's equation. It was observed that, after anneal, the tunnel region becomes slightly wider due to the B diffusion. The current–voltage ($I-V$) curves were calculated with a two-band model using the light electron mass of 0.19 m_0 and the indirect band gap of 1.12 eV. While the direct tunneling two-band model is not the correct physics for Si, it should predict the correct trends in the magnitude of the current for different junction potentials. These calculations show the current being reduced after RTA, which is consistent with the wider tunnel region, but is inconsistent with the experimental data, as will be shown.

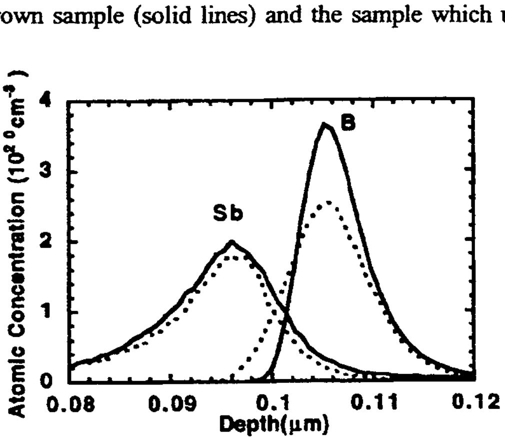


FIG. 2. SIMS atomic profiles of the Si RITD, as grown and after a post-growth anneal of 700 °C for 1 min. The region of the device, including the δ-doped layers and the intrinsic Si spacer, has been magnified. A linear scale was used to emphasize the effects of the anneal. The as-grown sample is shown with solid lines. The sample annealed at 700 °C is shown with dashed lines.

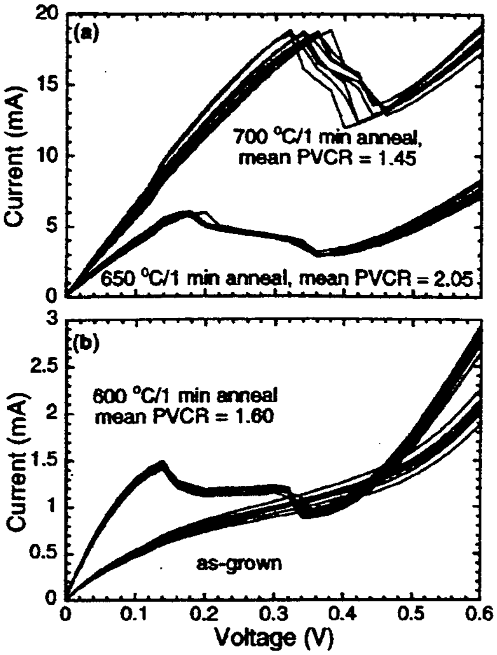


FIG. 3. $I-V$ characteristics for seven adjacent devices after the specified anneal. All measurements are on 18-μm-diam diodes and were taken at room temperature. The results are shown on two graphs because of the increased currents after the anneal.

The fabrication process for the RITDs has been previously described.[14] Room-temperature $I-V$ curves were measured from the top of the mesa to the backside Al contact. $I-V$ measurements for seven adjacent devices, having a diameter of 18 μm, for the as-grown samples and for each anneal condition, are presented in Fig. 3. It is seen that the as-grown RITDs do not exhibit NDR, but have $I-V$ characteristics of backwards diodes. NDR was observed following a postgrowth RTA between 600 and 700 °C. In the following discussion we focus on four characteristics of these curves: PVCR, the voltage V_p at which the relative maximum in the current I_p occurred, and the peak current density J_p.

The average PVCR was 1.60, 2.05, and 1.45 after a 1 min RTA of 600, 650, and 700 °C, respectively. These room-temperature values are significantly larger than reported PVCR values for epitaxial Si tunnel diodes at any temperature.[5–14,17] It is observed that V_p shifted to larger values with higher anneal temperature. The plot of V_p vs I_p was linear. The intrinsic series resistance obtained from the slope of the curve had a value of 12.2 Ω. The intrinsic peak voltage of the diodes was obtained from the intercept and had a value of 117 mV. This agreed well with the predicted value of 90 mV from the model. The peak current density increased with the anneal temperature from 5.7×10^2 A/cm^2 after the 600 °C/1 min anneal to 7.3×10^3 A/cm^2 after the 700 °C/1 min anneal. It is not clear why the device characteristics improved with anneal temperature. The diffusion of the B, observed by SIMS after the 700 °C RTA, widened the tunnel junction by compensating the *n*-type dopants and should result in a lower peak tunnel current. The simplest explanation for device improvement is the reduction of defects and increased activation of dopants in the quantum wells by the anneal. In the band-structure diagram we assumed complete activation at the growth temperature based on prior work on highly doped thick layers. This assumption may not be valid for the δ-doped layers and will be investigated further.

In summary, we have reported on the growth and characterization of Si RITD, which have significantly improved PVCR (>2) compared to previously reported epitaxial Si diodes. The growth was accomplished using low-temperature (320 °C) epitaxy to minimize the segregation of the dopants, which, in turn, permitted a very high concentration of *p* and *n* dopants in the quantum wells. A postgrowth RTA between 600 and 700 °C was required to observe NDR. The low growth and anneal temperatures and the thin epitaxial thickness requirement make these device suitable for CMOS and HBT integration.

This work was supported by the Office of Naval Research, National Science Foundation, and Defense Advance Research Projects Agency. The authors thank Lt. Col. Gernot Pomrenke for his encouragement throughout the project.

[1] U. Konig, Phys. Scr. **68**, 90 (1996).
[2] D. J. Paul, Thin Solid Films **321**, 172 (1998).
[3] A. Seabaugh, X. Deng, T. Blake, B. Brar, T. Broekaert, R. Lake, F. Morris, and G. Frazier, Tech. Dig. Int. Electron Devices Meet. **98**, 429 (1998).
[4] L. Esaki, Phys. Rev. **109**, 603 (1958).
[5] H. C. Liu, D. Landheer, M. Buchanan, and D. C. Houghton, Appl. Phys. Lett. **52**, 1809 (1988).
[6] S. S. Rhee, J. S. Park, R. P. G. Karunasiri, Q. Ye, and K. L. Wang, Appl. Phys. Lett. **53**, 204 (1988).
[7] E. Kasper, H. Dambkes, and J.-F. Luy, Tech. Dig. Int. Electron Devices Meet. **88**, 558 (1988).
[8] U. Konig, M. Kuisl, J.-F. Luy, and F. Schaffler, Electron. Lett. **25**, 1169 (1989).
[9] U. Gennser, V. P. Kesan, S. S. Iyer, T. J. Bucelot, and E. S. Yang, J. Vac. Sci. Technol. B **8**, 210 (1990).
[10] A. Zaslavsky, D. A. Grutzmacher, Y. H. Lee, W. Ziegler, and T. O. Sedgwick, Appl. Phys. Lett. **61**, 2872 (1992).
[11] K. Ismail, B. S. Myerson, and P. J. Wang, Appl. Phys. Lett. **59**, 973 (1991).
[12] Z. Matutinovic-Krstelj, C. W. Liu, X. Xiao, and J. C. Sturm, Appl. Phys. Lett. **62**, 603 (1993).
[13] M. R. Sardela, Jr., H. H. Radamson, and G. V. Hansson, Appl. Phys. Lett. **64**, 1711 (1994).
[14] S. L. Rommel, T. E. Dillon, M. W. Dashiell, H. Feng, J. Kolodzey, P. R. Berger, P. E. Thompson, K. D. Hobart, R. Lake, A. C. Seabaugh, G. Klimeck, and D. Banks, Appl. Phys. Lett. **73**, 2191 (1998).
[15] S. Rommel, T. Dillon, P. Berger, R. Lake, P. E. Thompson, K. D. Hobart, A. C. Seabaugh, and D. S. Simons, Tech. Dig. Int. Electron Devices Meet. **98**, 1035 (1998).
[16] M. Sweeny and J. Xu, Appl. Phys. Lett. **54**, 546 (1989).
[17] S. L. Rommel, T. E. Dillon, P. Berger, P. E. Thompson, K. D. Hobart, R. Lake, and A. C. Seabaugh, Electron. Device Lett. **20**, 329 (1999).
[18] E. D. Richmond, J. G. Pellegrino, M. E. Twigg, S. Qadri, and M. T. Duffy, Thin Solid Films **192**, 287 (1990).
[19] P. E. Thompson, M. E. Twigg, D. J. Godbey, K. D. Hobart, and D. S. Simons, J. Vac. Sci. Technol. B **11**, 1077 (1993).

Phonon Assisted Tunneling in Si / SiGe Tunnel Diodes

R. Lake

Raytheon Systems, PO 660246, MS35, Dallas, TX 75266

(December 21, 1999)

Abstract

We have derived an expression for the deformation potential phonon-assisted tunneling current in a Si / SiGe tunnel diode. We use a localized orbital band-model within the non-equilibrium Green function formalism. The non-equilibrium Green function approach allows us to model the 2D-2D tunneling from the quantum states that form in the δ-doped wells of the contacts. It also allows us to investigate the effects of incoherent broadening of those 2-D states due to scattering from the extremely high concentration of ionized dopants in those regions.

I. SI / SIGE INTERBAND PHONON ASSISTED TUNNEL THEORY

We derived the following expression for the deformation potential phonon assisted tunneling current in a Si / SiGe tunnel diode.

$$\begin{aligned} J_1 = \frac{e(D_t K)^2}{2\rho\omega\Delta} \int \frac{d^2\mathbf{k}_{\mathcal{L}}}{4\pi^2} \int \frac{d^2\mathbf{k}_{\mathcal{R}}}{4\pi^2} \int \frac{dE}{2\pi} \sum_{L=1}^{N} \mathrm{tr} \Big\{ & G^A_{L,1}(\mathbf{k}_{\mathcal{L}}, E) \Gamma^{\mathcal{L}}_{1,1}(\mathbf{k}_{\mathcal{L}}, E) G^R_{1,L}(\mathbf{k}_{\mathcal{L}}, E) \\ & \Big[f_{\mathcal{L}}(E) \left(1 - f_{\mathcal{R}}(E - \hbar\omega)\right) A^{\mathcal{R}}_L(\mathbf{k}_{\mathcal{R}}, E - \hbar\omega)(n_B(\hbar\omega) + 1) \\ & + f_{\mathcal{L}}(E) \left(1 - f_{\mathcal{R}}(E + \hbar\omega)\right) A^{\mathcal{R}}_L(\mathbf{k}_{\mathcal{R}}, E + \hbar\omega) n_B(\hbar\omega) \\ & - \left(1 - f_{\mathcal{L}}(E)\right) f_{\mathcal{R}}(E - \hbar\omega) A^{\mathcal{R}}_L(\mathbf{k}_{\mathcal{R}}, E - \hbar\omega) n_B(\hbar\omega) \\ & - \left(1 - f_{\mathcal{L}}(E)\right) f_{\mathcal{R}}(E + \hbar\omega) A^{\mathcal{R}}_L(\mathbf{k}_{\mathcal{R}}, E + \hbar\omega)(n_B(\hbar\omega) + 1) \Big] \Big\} \end{aligned} \tag{1}$$

where $A^{\mathcal{R}}_L$ is the component of the spectral function injected from the right contact.

the value to be 5.6e8 eV/cm which is a factor of 2.28 larger than the deformation potential for the TA phonons. We note that for the X-X intervalley phonon couplings, the TO phonon deformation potential of 2.0×10^8 eV/cm is approximately 7 times larger than the TA phonon deformation potential of 0.3×10^8 eV/cm [4].

We now discuss the integrals in Eq. (1) specifically for the case of the Si Esaki tunneling problem assuming that the left contact is n-type and the right contact is p-type. The electrons that contribute to the phonon-assisted tunneling are in the 4 transverse X valleys since they have the light mass in the tunneling direction. Therfore, $\mathbf{k}_{\mathcal{L}}$ is centered at the conduction band minimum, $\mathbf{k}_0 \approx 0.85\pi/\Delta(100)$, and the integral $\int \frac{d^2\mathbf{k}_{\mathcal{L}}}{4\pi^2}$ is over the ellipse in the $x-y$ plane centered around $\mathbf{k}_0$. All quantities with a subscript or superscript $\mathcal{L}$ are evaluated in the (100) electron X valley. For the numerical calculation, the integral $\int d^2\mathbf{k}_{\mathcal{L}}$ is approximated as

$$\pi \int_{k_y=0,k_x=\mathbf{k}_0}^{k_x=k_x^{\max}} dk_x\, k_x \; + \; \pi \int_{k_x=\mathbf{k}_0,k_y=0}^{k_y=k_y^{\max}} dk_y\, k_y \tag{3}$$

Eq. (1) is multiplied by a factor of 4 to account for the 4 equivalent valleys.

The right contact is p-type. Therefore, all quantities with a subscript or superscript $\mathcal{R}$ are evaluated in the hole Γ valley centered at (000). The integral $\int d^2\mathbf{k}_{\mathcal{R}}$ is approximated as

$$2\pi \int_0^{k_\Gamma^{\max}} dk_\Gamma k_\Gamma \tag{4}$$

Both integrals (3) and (4) occuring in Eq. (1) are discretized and evaluvated using trapezoidal rule.

II. RECURSIVE GREEN FUNCTION ALGORITHM

The Green function elements appearing in Eq. (1) are obtained using the recursive Green function algorithm described in reference [2] and briefly outlined below. To obtain the first block row of the Green function, $G^R_{1,L}$, we first calculate the right connected (or surface) Green function, $g^r_{L,L}$, from

$$g^r_{L,L} \;=\; \left[E - D_L - t_{L,L+1} g^r_{L+1,L+1} t_{L+1,L}\right]^{-1} \quad . \tag{5}$$

The exact diagonal block $G^R_{1,1}$ is then obtained from

$$G^R_{1,1} \;=\; \left[E - D_1 - t_{1,0} g^\ell_{0,0} t_{0,1} - t_{1,2} g^r_{2,2} t_{2,1}\right]^{-1} \tag{6}$$

where $g^\ell_{0,0}$ is the bulk surface Green function on the left. The first block row is evaluated using

$$G^R_{1,L} = G^R_{1,L-1}(-t_{L-1,L}) g^r_{L,L} \quad . \tag{7}$$

The advanced Green function is obtained using the relation $G^A_{L,1} = \left[G^R_{1,L}\right]^\dagger$.

To obtain the last column, $G^R_{L,N}$, the left connected Green function, $g^\ell_{L,L}$, is calculated using the equivalent of Eq. (5) except starting with $g^\ell_{0,0}$,

$$g^\ell_{L,L} \;=\; \left[E - D_L - t_{L,L-1} g^\ell_{L-1,L-1} t_{L-1,L}\right]^{-1} \quad . \tag{8}$$

Then the exact $G_{N,N}$ is found using the equivalent of Eq. (6).

$$G^R_{N,N} \;=\; \left[E - D_N - t_{N,N-1} g^\ell_{N-1,N-1} t_{N-1,N} - t_{N,N+1} g^r_{N+1,N+1} t_{N+1,N}\right]^{-1} \tag{9}$$

In Eq. (9) the quantity $t_{N,N+1} g^r_{N+1,N+1} t_{N+1,N}$ is the boundary self energy that has been brought in from the right. Similarly, in Eq. (6), the quantity $t_{1,0} g^\ell_{0,0} t_{0,1}$ is the boundary self energy that has been brought in from the left. Finally, the last column is generated from

$$G^R_{L,N} = g^\ell_{L,L}(-t_{L,L+1}) G^R_{L+1,N} \quad . \tag{10}$$

REFERENCES

[1] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, J. Phys. C: Solid State Physics **5**, 21 (1972).

[2] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).

[3] R. A. Logan and A. G. Chynoweth, Phys. Rev. **131**, 89 (1963).

[4] C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).

Band offset measurement of the ZnS/Si(001) heterojunction

B. Brar, R. Steinhoff, A. Seabaugh, X. Zhou,* S. Jiang,* and W. P. Kirk*

Raytheon TI Systems, Dallas, TX 75243.
*NanoFAB Center, Texas A&M University, College Station, TX 77843-4242
EMail: brar@resbld.csc.ti.com

Abstract. High-quality ZnS layers on silicon have recently been realized by initiating MBE growth on a vicinal Si (001) surface that has been terminated with a single monolayer (ML) of As. We report the first measurement of the electronic transport properties of the ZnS/As(1ML)/n-Si(001) heterostructure. Temperature-dependent current-voltage measurements show that the transport is characterized by an activation energy of 1.00±0.04 eV for the Si/ZnS conduction band offset. A similar activation energy of 1.02±0.04 eV is obtained for the opposite bias polarity, corresponding to transport over the Al/ZnS Schottky barrier.

1. Introduction

The realization of high-quality compound semiconductors that are lattice-matched to a silicon (Si) substrate will promote the development of band-gap-engineered devices in Si technology. One candidate for a barrier material is zinc sulfide (ZnS), which has an energy band gap of 3.6 eV and a zinc-blende (cubic) crystal structure with a lattice constant of 5.42 Å. By comparison, silicon has a bandgap of 1.12 eV and a diamond (cubic) crystal structure with a lattice constant of 5.431 Å. The band offset of a molecular beam epitaxy (MBE) grown ZnS/Si(111) heterojunction was measured by Maierhofer et al. [1], using photoelectron spectroscopy. They reported a valence-band offset of ΔE_V = 0.7 eV and a conduction-band offset of ΔE_C= 1.7 eV. The large band offsets in a straddled band line-up suggest that ZnS is an excellent choice for use as an insulating barrier that is lattice-matched to silicon. Recently, high- quality ZnS layers on silicon have been realized by initiating MBE growth on a vicinal Si (001) surface that has been terminated with a single monolayer (ML) of As [2]. In the present paper, we report the first electronic transport properties of the ZnS/As(1 ML)/n-Si(001) heterostructure.

2. Growth and Experiment

The samples were grown in a VG80S MBE chamber on phosphour-doped (1-10 Ω-cm) vicinal Si substrates (100 surface with a 4° off-cut towards $[0\bar{1}1]$). Details of the ZnS growth can be found in reference [2]. Briefly, growth was initiated with a 200 nm thick not-intentionally doped silicon buffer layer at a growth temperature of 620 ° C. The sample was then annealed at 850 °C to produce a double-stepped Si surface, as confirmed by reflection high-energy electron diffraction. After the substrate was cooled to room temperature, more than 1 ML of As was deposited. The excess (beyond 1 ML) As was then removed with a 600 °C anneal. Finally, 200 nm of ZnS was grown from a solid ZnS source at a substrate temperature of 50 °C followed by a 10 minute anneal at 320 °C. Both Al and Au contacts were e-beam evaporated and patterned on the sample. Metal covered the entire wafer surface except a 10 μm

CCC Code 0-7803-3883-9/98/$10.00 © 1998 IEEE

spacer region, which isolated 15, 150, and 1500 µm diameter circular devices. In the case of Al, the rings were etched in Al-leach; in the case of Au, a lift-off process was used to define the device. The large-area greater surface outside of these circles was grounded in the electrical measurements.

3. Results and Discussion

Figure 1 is a plot of the *J-V* characteristics of the 150 µm diameter device for selected temperatures. The bias was swept in two different directions (arrows indicate the sweep direction). The magnitude of the voltage was swept from 0 to 2 to 0 V at a rate of ~1 V/min. The hysteresis predominant at lower temperatures is presumably related to the charging time of slow traps in the ZnS barrier. The voltage dependence of the current at fixed temperature indicates a nearly symmetric device. The non-exponential dependence on applied bias (large non-ideality factor) is typical for a heterostructure barrier that does not change with applied voltage. Figure 2 is an Arrhenius plot constructed from the forward-sweep data in Fig. 1 to determine the activation energy E_c for the conduction mechanism. At lower temperatures, where the curves in Fig. 2 flatten out, the current measured is presumably through defect-assisted transport (e.g., via grain boundaries or traps in the ZnS) with a relatively small thermal activation energy. At higher temperatures, the data indicate a thermally-activated process with a well-defined activation energy. The activation energy is obtained at each bias point by fitting the data to a conventional thermionic emission model $J = J_0 \bullet T^2 \exp(-E_{act}/kT)$ [3], and plotted in Fig. 3 as a function of the applied bias. Note that an activation energy obtained from the Arrhenius plot is the extrapolated value at zero temperature, and is only accurate to within $kT \approx 40$ meV. Since the detail at low bias in Fig. 3 is smaller than this, it is not analyzed.

The activation energy may be related to the band offset by determining the chemical potential as a function of voltage and temperature. For a metal (with $E_F \gg kT$), the chemical potential is independent of both of these parameters, and the value of 1.02±0.04 eV obtained from Fig. 3 for negative biases, *is*

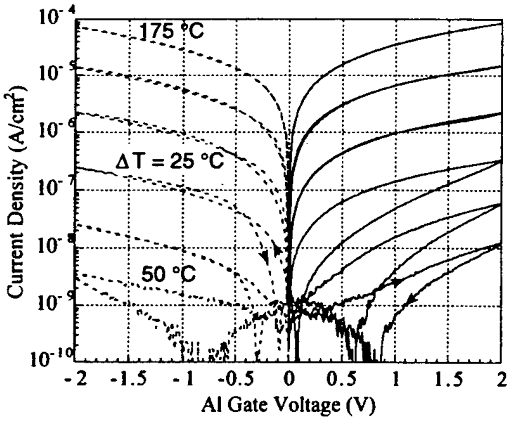


Fig. 1. Temperature dependence of the current densities through Al/ZnS/As(1ML)/Si diodes. The temperature is varied from 50 °C to 175 °C in 25 °C steps.

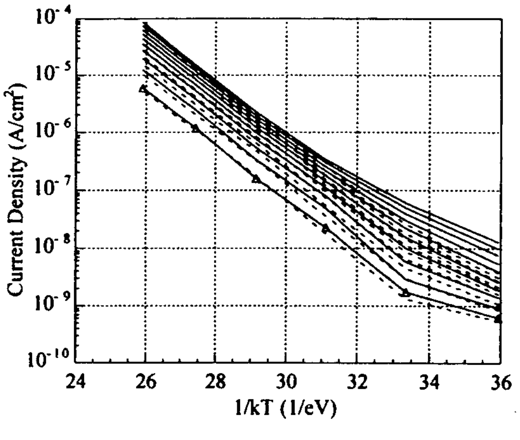


Fig. 2. I-V Arrhenius plot to extract activation energies in Al/ZnS/As(1ML)/Si diodes. Solid lines are positive gate biases (from 0.2 V to 2.0 V in 0.2 V steps) and the dashed lines are negative gate biases.

the Al/ZnS Schottky barrier height. For a non-degenerate semiconductor with fully-ionized donors, however, charge neutrality in the bulk forces $(E_C - E_F)/kT$ to remain constant as the temperature changes, and the extrapolated E_F at T = 0 is approximately E_C. Therefore, the activation energy of 1.00 eV measured for small positive gate voltages is approximately the conduction band offset ΔE_C. This value is considerably lower than the value of 1.7±0.2 eV obtained by photoelectron spectroscopy between ZnS and non-As-terminated Si(111) [1]. This difference in the conduction band offset is partly due to the substrate orientation. Also, some of the difference may be due to a modification of the interface dipole charge caused by the As monolayer.

To understand the bias (in)dependence of the extracted activation energy we performed capacitance-voltage (*C-V*) measurements on the above devices. The *C-V* data were noisy and hysteretic. Also the data for the devices characterized above showed no evidence of electron accumulation at the ZnS/Si interface. The hysteresis in the *C-V* signified field-driven transfer of charge within the ZnS, i.e. mobile ions or trapped electrons in the ZnS. Electron traps in the ZnS would account for the lack of image-force barrier lowering in Fig. 3, which should be roughly 50 meV when 2 V is applied across the ZnS. If traps were in the ZnS, and the ZnS charged negatively when current flowed, then the electric field at the ZnS/Si interface (causing barrier reduction) would have been smaller than the trap-free case. Our inability to measure an accumulation capacitance in these devices is consistent with our hypothesis that there is no significant electric field at the ZnS/Si interface even under bias (shown schematically in Fig. 3).

To confirm that the activation energy represents an interface property and not a bulk transport property of the ZnS, additional metal/ZnS/As(1ML)/Si devices were fabricated in an identical manner, but with Au electrodes. Previous work has determined the Au/ZnS Schottky barrier height to be 1.2 eV larger than the Al/ZnS Schottky barrier height [4]. Temperature dependent measurements on these Au/ZnS/ As(1ML)/Si-n showed that, consistent with our interface-barrier-limited current model, a large (~20x) reduction in current was observed as plotted in Fig. 4 (only forward sweep data are shown). The asymmetry in the *J-V* characteristic is also consistent with our model of Schottky emission. A well-

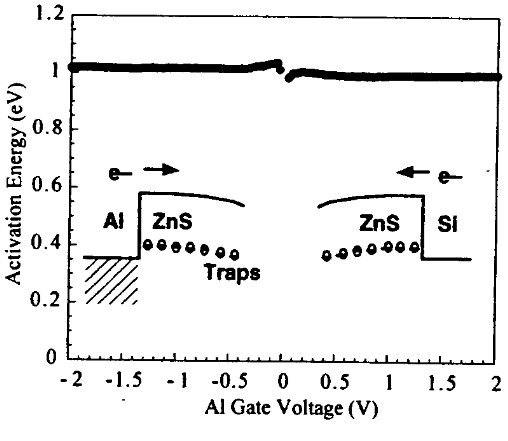


Fig. 3. Extracted activation energies in Al/ZnS/As(1ML)/Si diodes.

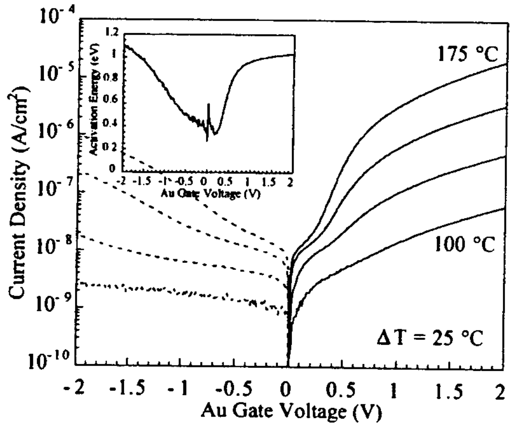


Fig. 4. Temperature dependence of the current densities through Au/ZnS/As(1ML)/Si diode. The inset shows the extracted activation energies as a function of bias.

defined activation energy (see inset of Fig. 4) exists only for positive gate voltages greater than approximately 0.6 V where it starts to level-off around 1.0 eV. This value for the ZnS/Si barrier height agrees with the value for devices with Al electrodes. For negative biases, defect transport processes dominate any bulk emission over the large Au/ZnS barrier height. Figure 4 shows that these defect processes do not have a single well-defined activation energy.

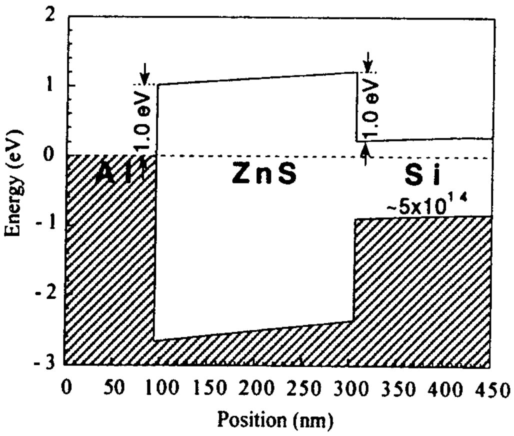


Fig. 5. Zero-bias energy band diagram for Al/ZnS/As(1ML)/Si structure.

4. Conclusions

The Al/ZnS/As(1ML)/Si-n heterostructure was found to be characterized by a ZnS/Si conduction band offset ΔE_C = 1.00±.04 eV, and a Al/ZnS Schottky barrier height of 1.02±0.04. *C-V* measurements, *I-V* hysteresis, and the lack of a reduction in the barrier height due to image force lowering (expected to be ≈ 50 meV at a bias of 2 V) suggests that there are traps in the ZnS that charge negatively under bias. Figure 5 is a schematic of the proposed energy band diagram for the Al/ZnS/As(1ML)/Si-n heterostructure. The approximately 1 eV ZnS barrier heights to Al and Si, being much larger than *kT* at room temperature, are encouraging for the use of ZnS as an insulating epitaxial barrier for Si devices.

Acknowledgments

The authors gratefully acknowledge several stimulating discussions with Drs. R. Lake, G. Wilk, and T. Moise, and the excellent technical support of P. Stickney. This research has been supported in part by DARPA Contract No. F49620-96-C-0006, AFOSR Grant No. F49620-96-1-0242, and NSF Grant No. ECS-9306293.

References

[1] C. Maierhofer, S. Kulkarni, M. Alonso, T. Reich, and K. Horn, *J. Vac. Sci. Techn.*, vol. B9, pp. 2238, 1991.

[2] L. T. Romano, R. D. Bringans, X. Zhou, and W. P. Kirk, *Phys. Rev. B*, vol. 52, pp. 11202, 1995; and X. Zhou, S.Jiang, and W.P. Kirk, *J. Appl. Phys*, vol. 82 (1997).

[3] S. M. Sze, *Physics of Semiconductor Devices*. 2nd Ed. New York: John Wiley & Sons, 1981.

[4] M. Aven and C. A. Mead, *Appl. Phys. Lett*, vol. 7, 1965.

Appendix 16

SiO_2 film thickness metrology by x-ray photoelectron spectroscopy

Z. H. Lu and J. P. McCaffrey
Institute for Microstructural Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

B. Brar, G. D. Wilk, and R. M. Wallace
Corporate Research Laboratories, Texas Instruments, Inc., Dallas, Texas 75243

L. C. Feldman[a]
Bell Laboratories, Lucent Technology, Murry Hill, New Jersey 07974-0636

S. P. Tay[b]
Telecom Microelectronics Centre, Northern Telecom Ltd., Nepean, Ontario K2H 8V4, Canada

(Received 24 July 1997; accepted for publication 9 September 1997)

Silicon dioxide films grown by industrial thermal furnace, rapid thermal, and low-pressure thermal methods were measured by x-ray photoelectron spectroscopy, transmission electron microscopy (TEM), spectroscopic ellipsometry, and capacitance–voltage analysis. Based on TEM measurements, the photoelectron effective attenuation lengths in the SiO_2 and Si are found to be 2.96 ± 0.19 and 2.11 ± 0.13 nm, respectively. The oxide physical thicknesses (range from 1.5 to 12.5 nm) as measured by all above techniques are in good agreement. The electrical thickness is noted to be slightly thicker than the physical thickness. © *1997 American Institute of Physics.* [S0003-6951(97)01845-7]

Ultrathin SiO_2 films are critical for novel nanoelectronic devices as well as for conventional deep submicron ultra large scale integrated (ULSI) circuit where the gate oxide is reduced to less than 3 nm. Precise thickness measurement of these ultrathin films is very critical in the development of future Si-based devices. Oxide thickness is commonly measured by ellipsometry,[1] but as film thicknesses are scaled down to several atomic layers, surface analytical techniques, such as x-ray photoelectron spectroscopy (XPS), become applicable tools to quantify these films. An XPS measurement offers the additional advantage of providing information such as surface contamination and chemical composition of the film. Thus an XPS system will be able to map out dielectric film physical thickness, chemical composition, and the nature of the chemical bonding. It is well known, however, that large discrepancies exist for the photoelectron effective attenuation length in SiO_2, where values from 2 to 4 nm have been reported and compared to theoretical predictions for the inelastic mean-free path.[2] This critical parameter must be resolved before XPS can become a standard method. In this article we describe oxide layer thickness measurements by XPS, transmission electron microscopy (TEM), spectroscopic ellipsometry, and capacitance–voltage ($C-V$). Correlation of these measurements allows us to extract a precise value of the mean-free path.

The methodology of XPS measurements is well described in Ref. 3. Briefly, the oxide film thickness d_{oxy} is determined by the Si $2p$ core level intensity ratio of the oxidized silicon film I_{oxy} and substrate silicon I_{Si} by

$$d_{oxy}=\lambda_{oxy}\sin\alpha\ \ln[I_{oxy}/(\beta I_{Si})+1],$$

where λ_{oxy} is the photoelectron effective attenuation length in the oxide film, α is a photoelectron take-off angle, and $\beta=I_{oxy,\infty}/I_{Si,\infty}$ measured on photoelectron opaque films ($\geqslant 15$ nm at $\alpha\geqslant 70°$). The above equation was derived based on the assumption that the photoelectron signal has an exponential depth distribution function. This assumption has been confirmed experimentally[4] and theoretically by a Monte Carlo simulation.[5] The effective attenuation length and inelastic scattering mean-free path are considered equivalent when elastic scattering effects.[5]

We used a PHI 5500 XPS system which is equipped with a monochromatic Al K_α source and a hemispherical electron energy analyzer. The largest available acceptance angle ($\sim 28°$) of the analyzer was used to reduce possible photoelectron forward scattering effects. The photoelectron take-off angle was fixed at 77.5° for all measurements. Selected samples were tested by sample rotation during measurement to verify that there were no forward scattering artifacts at the take-off angle used. It should be noted that care is required if the oxide thickness is measured at 45° and 35.26° take-off angles as major x-ray photoelectron diffraction peaks occur for the ⟨110⟩ and the ⟨111⟩ direction, respectively.[6] These diffraction peaks, however, can be used to calibrate the sample stage and the take-off angle precisely, by measuring the ⟨110⟩ and ⟨111⟩ forward scattering peaks using high angular resolution, which is $\sim 4°$ in our instrument. Also measurements at low ($<20°$) take-off angles should be done with extreme caution. At these extreme angles, photoelectron refraction, which can be as large as 2° at a 10° take-off angle,[6] and sample manipulator precision amount to a large error which can be as high as 40%. At high ($>25°$) take-off angles, however, extremely consistent measurements in film thicknesses (within 0.1 nm) have been observed. Measurements on thick thermal oxides and on H-terminated Si (100) samples yield $\beta=0.75$.

The last parameter to be determined is the effective attenuation length. In order to obtain λ_{oxy}, a set of samples with known thicknesses has to be measured. TEM is a technique that can directly measure the thickness without any assumption or invoking any other reference samples, and therefore TEM determined thickness was used to find λ_{oxy}.

---

[a] Present address: Dept. of Physics and Astronomy, Vanderbilt University, Nashville, TN.

[b] Present address: AG Associates, San Jose, CA.

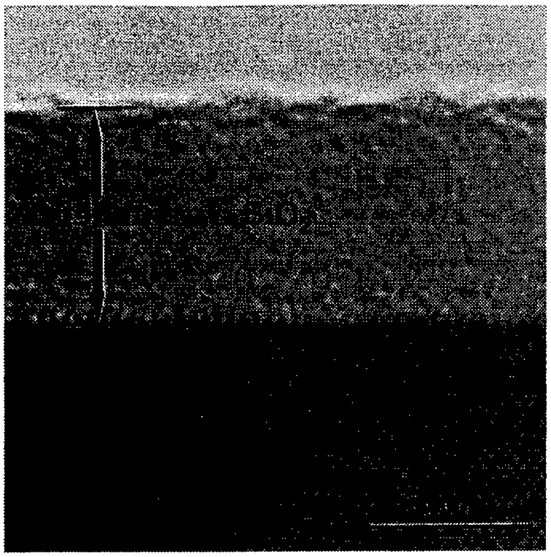


FIG. 1. A TEM micrograph showing a 6.4 nm SiO_2 film on Si (100).

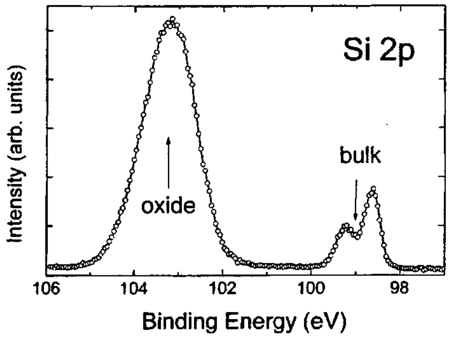


FIG. 2. An XPS Si $2p$ core level spectrum of a 6.4 nm oxide film.

Different types of samples were produced by several methods including standard industrial thermal furnace oxides (FO) (6 in. wafer), rapid thermal oxides (RTO) (6 in. wafer), and low-pressure thermal oxides (LPO) (4 in. wafer). The uniformity of an industrial furnace oxide is very high. The thickness variation across the wafer is less than 1.5% of the total thickness. The majority of TEM samples were made from furnace oxides. Two RTO samples were also used. The thickness variation for RTO is less than 2%. One LPO sample was also used.

The TEM measurements were carried out on a Phillips EM 430T system operating at 250 keV. Cross-sectional samples were made by a small-angle cleavage technique.[7] Figure 1 shows a TEM micrograph obtained on a 6.4 nm oxide sample. As the substrate lattice can be well-resolved in a TEM micrograph, the Si lattice spacing is used as a built-in reference length gauge to measure the oxide thickness. An uncertainty in a TEM measurement lies in the determination of the SiO_2/Si (100) interface and SiO_2 surface. One atomic layer variation is certainly possible due to the intrinsic two domain interface structure with each domain differing by one atomic step height. The nature of an amorphous-crystalline interface prevents oxide film thickness determination of less than 0.3 nm accuracy by TEM. Measurements on several samples from the same wafer indicated a 0.3 nm thickness variation, larger than the oxide uniformity across the wafer, as determined by ellipsometry which has a sampling area on the order of several mm^2. The sampling region of a given TEM image is 10–100 nm. Rather than performing multiple TEM measurements on a given wafer, we preferred to examine a wide range of oxide thicknesses to facilitate comparison amongst several techniques. The selected wafers provide us with not only statistical data but also a broad picture of measurement trend. In Table I we tabulate TEM measured thicknesses. The XPS data on the same set of samples are also shown in the table.

Figure 2 shows a Si $2p$ XPS spectrum recorded from the same sample as shown in Fig. 1. High energy resolution spectra were obtained by using 5.85 eV pass energy. The $2p$ spin-orbit splitting, $p_{3/2}$ and $p_{1/2}$, is clearly resolved for bulk substrate silicon. The $p_{3/2}$ position of 98.64 eV is in reference to the valence band maximum.[8] The spin-orbit splitting for the oxide film is not resolved due to the amorphous nature of the SiO_2 structure.[9] Since the two oxide and bulk peaks are well separated, I_{oxy} and I_{Si} can be easily determined by integrating the peak area after a Shirley background subtraction and the results also are listed in Table I. We found that the determination of the peak area by curve-fitting and direct area integration yield the identical result. Using the TEM and I_{oxy}/I_{Si} data, the photoelectron effective attenuation length can then be calculated (see Table I). Based on this set of data, the average photoelectron attenuation length for Si $2p$ core level in SiO_2 is found to be 2.96 nm with a standard deviation of 0.19 nm. It is noted that the two thinnest oxides have attenuation lengths lower than thicker oxides. This is attributed to the uncertainty of TEM in measuring ultrathin oxides. It has been suggested in Refs. 10 and 11 that thin oxides have attenuation lengths lower than those of thicker oxides, thus indicating a strained SiO_2 layer structure as proposed in Ref. 12. Recent studies[13–15] have shown that the experimental data in Ref. 12 originated from surface charging and core-hole relaxation. Furthermore, the XPS-measured thicknesses (range from ~1.5 to 12.5 nm) using 2.96 nm attenuation length agree well with thicknesses obtained by other techniques as will be discussed below.

TABLE I. Photoelectron attenuation length calculated based on TEM thickness and I_{oxy}/I_{Si} obtained from various samples.

Sample ID	nt1	D145	#2167	nt2	dt420198#24	nt3	nt4	D027	dt346537#18	Mean value (nm)
Oxidation	FO	RTO	LPO	FO	FO	FO	FO	RTO	FO	n.a.
TEM thickness (nm)	2.2	3.6	4.5	5.7	6.4	8.9	10.4	10.6	12.7	n.a.
I_{oxy}/I_{Si}	1.042	2.060	2.734	4.665	6.390	14.204	21.590	20.950	54.230	n.a.
λ_{oxy} (nm)	2.587	2.792	3.001	2.953	2.909	3.046	3.139	3.227	3.029	2.96±0.19
λ_{Si} (nm)	1.840	1.985	2.134	2.100	2.069	2.166	2.232	2.294	2.154	2.11±0.13

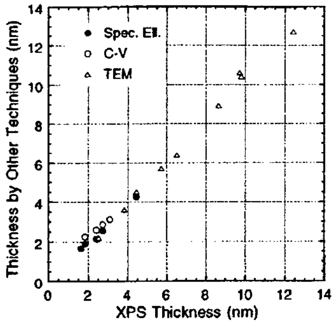


FIG. 3. XPS measured SiO_2 film thickness vs thickness measured by spectroscopic ellipsometry, $C-V$ analysis, and TEM.

A set of samples was also measured by spectroscopic ellipsometry and $C-V$ techniques. The details regarding the measurements can be found in Ref. 16. Figure 3 plots XPS thickness against thickness measurements by TEM, $C-V$, and spectroscopic ellipsometry (SE). Excellent agreement between TEM and SE measurements is observed. The oxide thickness derived from $C-V$ measurement, however, is found to be slightly higher than that obtained by other techniques. This is attributed to the electronic wave function that is peaked away from the oxide-silicon interface, thereby leading to a larger "electrical thickness."[16]

Thin Si-films-on-oxide is another important structure for quantum tunneling devices and for low-power consumption ULSI chips. Precise measurement of the silicon film thickness is also a critical issue. For a Si film on a thick, opaque oxide film (~a few nm depending on the photoelectron take-off angle), the thickness d_{Si} can be measured by

$$d_{Si}=\lambda_{Si}\sin\alpha\ \ln\{[(\beta I_{Si})/I_{oxy}]+1\},$$

where λ_{Si} is the photoelectron attenuation length in the Si film. It is established[3] that

$$\beta=(Y_{oxy}D_{oxy}\lambda_{oxy})/(Y_{Si}D_{Si}\lambda_{Si}),$$

where Y is the photoelectron yield (the percentage of electrons which escape without undergoing inelastic scattering), and D is the number of Si atoms per unit volume. In an inelastic scattering process where a photoelectron collides with a plasmon ω_p, the photoelectron loses kinetic energy of a defined plasmon energy $n\hbar\omega_p$ (where n is the number of plasmons generated).[17] The yield is obtained by dividing the area of the Si $2p$ peak by the sum of the areas of the $2p$ peak and all measurable plasmon loss peaks. We found $Y_{oxy}=0.69$ and $Y_{Si}=0.59$. The Si atomic density in the bulk Si and in the SiO_2 is 5.00×10^{22} and 2.28×10^{22} atoms cm^{-3}, respectively.[3] The photoelectron attenuation length in a crystalline Si can then be calculated. The calculated numbers are listed in Table I. The average λ_{Si} is found to be 2.11 nm with a standard deviation of 0.13 nm.

The published attenuation length in the oxide varies from 2 to 4 nm.[2] Our results for an average attenuation length of 2.96 nm falls in the middle of the range of values reported.[2] Several issues must be critically considered on studies of thin oxide films done previously: (1) the use of a nonmonochromatic photon source, (2) limited reference TEM samples covering a relatively narrow range of oxide thicknesses, (3) the use of single-wavelength ellipsometry reference samples, and (4) the use of techniques dependent on film density assumptions. The use of nonmonochromatic sources for XPS measurements results in the superposition of secondary lines and therefore leads to difficulties in the quantitative determination of peak area. Moreover, the abundance of stray electrons and Bremsstrahlung in the acquired spectra make quantitative analysis all the more difficult. Most of the studies using TEM reference samples examine a small number of samples with oxide thicknesses in a very narrow range. Our results show that substantial variability exists in the effective attenuation length derived from a broad selection of TEM measured thicknesses. Many earlier studies include "standards" derived from complimentary optical techniques such as ellipsometry. It is known however that single-wavelength ellipsometry has validity only in a thicker oxide film regime[18] and the use of this technique inevitably results in an overestimation of the attenuation length. Lastly, the use of techniques such as nuclear reaction analysis requires ultimately an assumption of the film density to derive a length scale from the data. The assumption of bulk density for an ultrathin film is problematic.

In summary, the Al $K\alpha$ XPS photoelectron mean-free path has been measured based on a large number of reference TEM samples. The average Si $2p$ photoelectron inelastic mean-free path in bulk Si and in the oxide film is found to be 2.11 ± 0.13 and 2.96 ± 0.19 nm, respectively. The oxide film physical thicknesses measured by TEM, XPS, and SE are now in excellent agreement with each other. Because of the charge carrier wave function spatial distribution, the oxide electrical thickness as measured by $C-V$ analysis is found to be slightly larger than the physical thickness.

[1] A. Kalnitshy, S. P. Tay, J. P. Ellul, S. Chongsawangvirod, J. W. Andrews, and E. A. Irene, J. Electrochem. Soc. **137**, 235 (1990).

[2] S. Tanuma, C. J. Powell, and D. R. Penn, Surf. Interface Anal. **17**, 927 (1991), and references therein.

[3] M. F. Hochella, Jr. and A. H. Carim, Surf. Sci. **197**, L260 (1988).

[4] D. F. Mitchell, K. B. Clark, J. A. Bardwell, W. N. Lennard, G. R. Massoumi, and I. V. Mitchell, Surf. Interface Anal. **21**, 44 (1994).

[5] C. J. Powell, A. Jablonski, S. Tanuma, and D. R. Penn, J. Electron Spectrosc. Relat. Phenom. **68**, 605 (1994).

[6] See, for example, C. S. Fadley, Prog. Surf. Sci. **16**, 275 (1984); S. A. Chambers, Adv. Phys. **40**, 357 (1991).

[7] J. P. McCaffrey, Microsc. Res. Tech. **24**, 180 (1993).

[8] Z. H. Lu, J-M. Baribeau, and T. E. Jackman, Can. J. Phys. **70**, 799 (1992).

[9] Z. H. Lu, M. J. Graham, D. T. Jiang, and K. H. Tan, Appl. Phys. Lett. **63**, 2941 (1993).

[10] R. P. Vasquez and F. J. Grunthaner, Surf. Sci. **99**, 681 (1980).

[11] M. H. Hecht, F. J. Grunthaner, P. Pianetta, L. I. Johansson, and I. Lindau, J. Vac. Sci. Technol. A **2**, 584 (1984).

[12] F. J. Grunthaner, P. J. Grunthaner, R. P. Vasquez, B. F. Lewis, J. Maserjian, and A. Madhukar, Phys. Rev. Lett. **43**, 1683 (1979).

[13] A. Pasquarello, M. Hybertsen, and R. Car, Phys. Rev. B **53**, 10 942 (1996).

[14] Y. Tao, Z. H. Lu, M. J. Graham, and S. P. Tay, J. Vac. Sci. Technol. B **12**, 2500 (1994).

[15] For a recent review, see Z. H. Lu, in *Fundamental Aspects of Ultranthin Dielectrics on Si-based Devices*, NATO ASI Series, edited by E. Garfunkel, E. Gusev, and A. Vul (Kluwer, Dordrecht, The Netherlands) (in press).

[16] B. Brar, G. D. Wilk, and A. C. Seabaugh, Appl. Phys. Lett. **69**, 2728 (1996).

[17] See, for example, L. C. Feldman and J. W. Mayer, *Fundamentals of Surface and Thin Film Analysis* (North-Holland, New York, 1986).

[18] S. Y. Kim and E. A. Irene, Rev. Sci. Instrum. **66**, 5277 (1995).

Appendix 17

IEDM, December 8, 1998, San Francisco

Transistors and Tunnel Diodes For Analog/Mixed-Signal Circuits and Embedded Memory

(Invited)

A. Seabaugh, X. Deng†, T. Blake†, B. Brar, T. Broekaert, R. Lake, F. Morris, and G. Frazier

Raytheon Systems Company, †Texas Instruments Incorporated
Dallas, TX 75243

Abstract

An integrated tunnel diode/transistor process can be used to increase the speed of signal processing circuitry or reduce power at the same speed; in memory applications, tunnel diodes can be used to reduce static power dissipation (>20× in Si, >1000× in III-V materials) relative to conventional approaches. This paper summarizes recent progress in InP and Si-based tunnel diodes[1] and circuits.

Introduction

Integrated circuit (IC) performance is limited by the finite gain-bandwidth product of the transistor technology. By adding a high speed tunnel diode (TD) to an existing transistor process, circuit speeds and power scaling can typically be improved. This is due to the high speed of the TD compared to the transistor, and the reduced component count of the TD/transistor circuit compared to the transistor circuit.

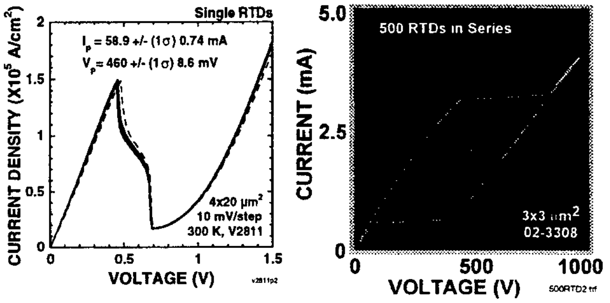
While Si TD/transistor technology was explored extensively in the 1960s and 70s, it suffered from being fundamentally discrete; the early "demise of the tunnel diode was signaled by the arrival in force of the integrated circuit" (1). Today, with the precise growth control of molecular beam epitaxy (MBE) the impediments to adding tunnel diodes to transistor processes are largely removed and large-scale TD/transistor ICs are being developed for the first time (2-4).

In this paper we show that the TD, in particular the resonant tunneling diode (RTD), can be produced with high uniformity and good reproducibility. RTDs been combined with the InP-based HEMT (high electron mobility transistor) in a high speed large scale integration (LSI) process to produce gigahertz clocked circuits (2,3) and recent progress in this technology will be summarized. The focus of this paper however will be to revisit the Si TD/transistor prospects, particularly the addition of the TD with CMOS. For this we use SPICE simulation, based on Texas Instruments' (TI) 0.25 μm CMOS models and an InP RTD model, fit to both dc and S-parameter measurements (4).

InP-Based RTD/HEMT Technology

The record highest speed transistors (HEMT), and TDs, (specifically, RTDs) have both been achieved on InP substrates with a cut-off frequency of 340 GHz (5) and switching speed of 1.5 ps (6) respectively. These devices have been combined with capacitors, resistors, and Schottky diodes in a monolithic IC format for gigahertz analog/mixed signal circuits and used to demonstrate: 4-bit, 3 GHz analog-to-digital converters (4), 3 GHz (50 dB spur free dynamic range) clocked quantizers (4), 3 GHz sample & hold circuits (55 dB linearity) (4), 5 GHz clock circuits (3), 2.5 GHz shift registers (3), and ultralow power SRAM (50 nW/bit) (7). Test equipment has limited the speeds at which these circuits could be characterized; clock rates in the range 10-100 GHz are projected from SPICE simulations.

These circuit demonstrations are made possible by the high uniformity of the MBE-grown RTD. This uniformity is shown in Fig. 1(a) where the current-voltage (I-V) characteristic of nine diodes are shown which span a 50 mm InP wafer. The dashed curves corresponding to the two outer RTD cells, closest to the wafer edge, while the solid lines are for the interior cells.



(a) (b)

Fig. 1: Resonant tunneling diodes (RTDs): (a) position dependence of current-voltage characteristic across a 50 mm InP wafer (the edge devices are indicated by a dashed line) and (b) five hundred series-connected InP-based RTDs showing ~3 % variation of peak and valley currents over the array. Both diodes use AlAs/InGaAsInAs/InGaAs/AlAs heterostructures.

---

[1] In this paper, "tunnel diode" refers broadly to p^+n^+ Esaki diodes and RTDs.

The RTD I-V characteristic is essentially determined by its layer dimensions, which set the diode peak current and peak voltage. Just as the turn-on voltage of a *p-n* diode is determined by the bandgap which is derived from the lattice spacing, the RTD critical parameters are also determined by bandgaps and lattice spacings. By MBE, thicknesses and compositions can be controlled precisely and uniformly as shown in Fig. 1 (a); even for an RTD with a peak current density exceeding 10^5 A/cm^2 (corresponding to a tunnel barrier thickness of approximately 6 monolayers) this layer dimension and the peak current are held essentially constant across the wafer.

Excellent local uniformity is shown in Fig. 1(b), a measurement of 500 RTDs connected in series. As the voltage is increased each of the 500 RTDs switches independently and in order from the lowest RTD peak current to the highest peak current. In this way the peak current of all 500 devices is sampled in a single measurement. On the return trace the switching order proceeds from the highest valley current to the lowest valley current and the valley current of all the devices is then also measured. From Fig. 1(b), the uniformity of peak and valley currents, which includes both material and process variations, can be seen to be less than ~3% in 500 devices.

Reproducibility is also under good control using MBE. Shown in Fig. 2 are measurements from two wafers which were grown consecutively then processed separately. The RTD peak current is plotted against device area showing that I-V characteristics have good run-to-run reproducibility in both growth and process.

CMOS/TD Circuits

A. *Memory Cells*

Tunnel diode growth on silicon is relatively immature. Recently however, CMOS-compatible interband tunnel diodes have been demonstrated (8,9) and show that a wide range of current-densities can be obtained spanning the range from 10^{-6} A/cm^2 (8) as required for embedded memory (10) to more than 20 kA/cm^2 (9) as needed for signal processing.

Dynamic random access memory (DRAM) cells have one access transistor and one capacitor in an integrated, highly specialized structure optimized for maximum density and minimum standby power. A process for embedding DRAM and logic on the same chip is desirable, but the periodic refreshing required for maintaining DRAM data adds complexity and consumes most of the DRAM power. A single-transistor (1T)/two-TD memory cell uses a DRAM cell topology, but with the addition of TDs eliminates the need for refreshing, and changes the cell storage from dynamic to static (SRAM) (7,10).

The tunnel SRAM cell is shown Fig. 3(a) where the currents I_3, I_2, and I_1 respectively represent the total leakage current through the access transistor and storage capacitor, the current through TD2, and the current through TD1. The leakage current versus node voltage and the I-V characteristics of the TDs are shown in Fig. 3(b). As shown in the same figure, the combined drive (I_2+I_3) and load (I_1) lines have two stable points, one near 1 V and the other near GND. It can be seen that the leakage current I_3 is completely compensated by the current difference I_1-I_2, thereby storing the logic level statically. (Note the TDs do not drive the bit line; data readout is DRAM-like using sense amplifiers).

In Table 1, using TI's 0.25 μm, 1.8 V, 0.7 μm metal pitch, CMOS logic-process models, the TD SRAM is compared against DRAM and SRAM in terms of layout area, relative areas, and estimated read/write cycle time and standby currents. The cases which were examined are, in order of Table 1: a 6T SRAM, a DRAM cell with a gate oxide capacitor (Cg), the TD SRAM (Fig. 3a) with Cg, a DRAM cell with a stacked-capacitor (Cst) cell instead of Cg, and the TD SRAM with Cst.

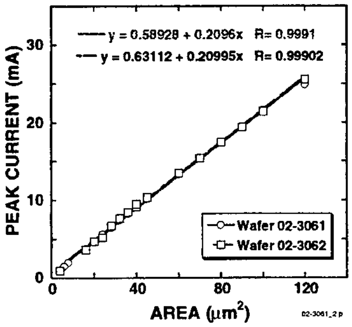
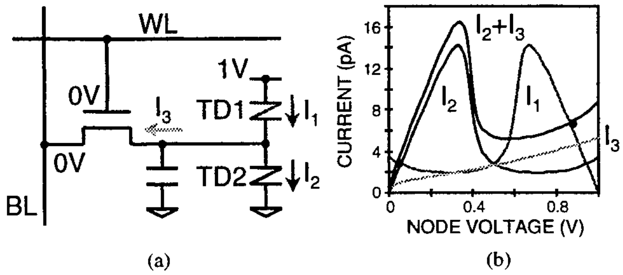


Fig. 2: Reproducibility of two separately grown and processed InP resonant tunneling diode wafers showing reproducibility and linearity of the diode formation process.



(a) (b)

Fig. 3: CMOS/TD SRAM cell: (a) schematic diagram and (b) current-voltage relations.

TABLE I
COMPARISON OF CMOS/TD SRAM CELLS WITH CMOS EMBEDDED DRAM AND 6T SRAM

Process	Cell	Area (μm^2)	Density Ratio (x6T)	Cycle Time (ns)	Standby Current (pA)
CMOS	6T SRAM	7.00	1.0	4	7
CMOS	1T+1Cg DRAM	5.79	1.2	18	31
CMOS +TD	1T+1Cg+2TD SRAM	7.02	1.0	18	3
CMOS +Cap	1T+1Cst DRAM	2.26	3.1	12	21
CMOS +Cap+TD	1T+1Cst+2TD SRAM	3.15	2.2	12	0.9

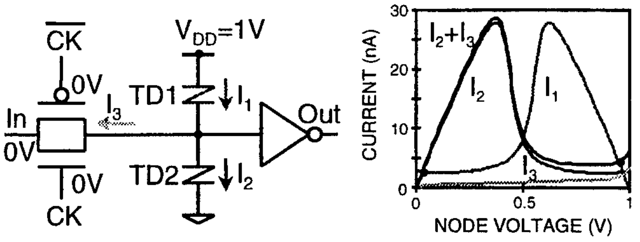
Standby current in Table 1 is due to MOSFET leakage in the 6T SRAM cells; in the TD SRAM cells leakage is the sum of MOSFET leakage and TD valley currents; in the DRAM cells standby current is mostly due to bit-line switching during refresh cycles, plus the leakage through the 0.5 µm width access transistor. By eliminating the refresh cycle, the embedded CMOS/TD SRAM achieves more than 20x power savings in comparison with embedded DRAM cell.

The TD SRAM cell size is less than half the size of the SRAM cell, with the TD minimum area set by the contact size. An additional mask level would be required to implement this TD SRAM, offset to some extent by a savings in chip area. As a comparison, a 256 Mbit DRAM cell, which uses lithography and patterning comparable to the designs here plus self-aligned contacts and other specialized features, is typically 0.8 μm^2 or less, but standby current, due to refresh, and due to the back-bias generator, can be several milliamperes. (Back bias is needed in triple-well DRAM arrays to increase the access transistors' threshold voltage and reduce leakage. Higher threshold voltages are not compatible with high-speed logic).

B. CMOS/TD Static Latches and Shift Registers

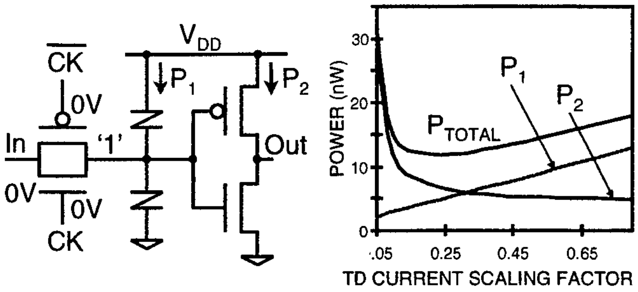
As in SRAM the TD pair also finds use in CMOS static latches, as shown in Fig. 4. To minimize static power consumption, the TD current is optimized, as outlined in Fig. 5. Too weak a TD current moves the stable point away from the supply rail resulting in an increase in the sub-threshold current of the output inverter; too strong a TD current, on the other hand, increases the standby current of the TD pair.

By cascading the latches in Fig. 4(a), a CMOS/TD static shift register (or D flip flop) is implemented as shown in Fig. 6(a). Here, a complementary pass gate is used to pass the data under the control of the clock signal.



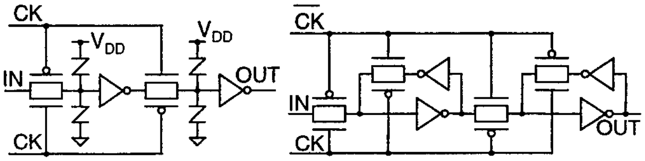
(a) (b)

Fig. 4: CMOS/TD static latch: (a) schematic diagram and (b) current-voltage relations.



(a) (b)

Fig. 5: Static power minimization: (a) schematic diagram and (b) power relationships.



(a) (b)

Fig.6: Static shift register comparison: (a) CMOS/TD and (b) CMOS only.

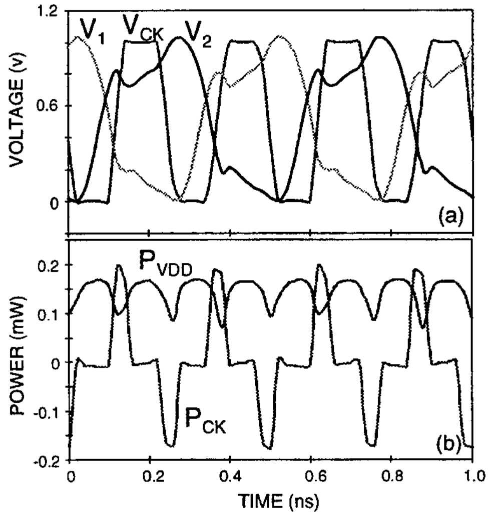
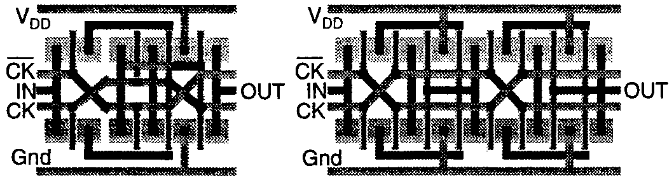


Fig. 7: CMOS/TD static shift register operating at the maximum clock frequency: (a) voltage waveforms and (b) power waveforms.

For comparison, a typical CMOS static shift register with a complementary pass gate is shown in Fig. 6(b). The feedback loop, consisting of an inverter and a complementary pass gate, is used to hold the storage voltage statically.

These circuits were simulated using the models discussed previously; interconnect parasitics were not included. The TD current was optimized to minimize static power consumption. Figure 7 shows the voltage and power waveforms of the CMOS/TD static shift register when operating at its maximum clock frequency. Since two inverters and two pass gates are eliminated, the CMOS/TD shift register has estimated 1.6× smaller area, 1.7× higher speed, and 2.4× lower power in comparison with its' CMOS counterpart, as shown in Fig. 8 and Table 2.

A single clock CMOS/TD static shift register was also similarly designed and compared with its CMOS counterpart, as shown in Fig. 9. The simulation shows that the CMOS/TD single-clock static shift register has estimated 3× smaller area, 2× higher speed, and 2× lower power in comparison with its CMOS counterpart, as shown in Table 3.

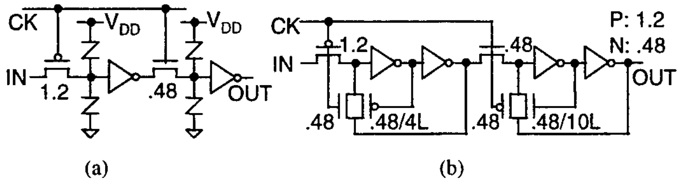


(a) (b)

Fig. 8: Symbolic layout comparison of (a) CMOS/TD and CMOS only static shift registers.

TABLE 2
COMPARISON OF CMOS AND CMOS/TD STATIC SHIFT REGISTERS

Process	VDD (V)	Vt (V)	Rel. Area	Max. Clock Rate (GHz)	Active Power (μW/GHz /bit)	Static Power (nW/bit)
CMOS	1.0	0.3	1.6×	2.38	50.0	10
CMOS/TD	1.0	0.3	1.0×	4.00	20.8	16
CMOS	0.5	0.15	1.6×	1.25	14.0	190
CMOS/TD	0.5	0.15	1.0×	2.27	5.7	215



(a) (b)

Fig. 9: Comparison of single-clock static shift registers: (a) CMOS/TD and (b) CMOS only.

TABLE 3
COMPARISON OF SINGLE-CLOCK CMOS AND CMOS/TD STATIC SHIFT REGISTERS

Process	VDD (V)	Relative Area	Max. Clock Rate (GHz)	Active Power (μW/GHz /bit)	Static Power (nW/bit)
CMOS	1.0	3×	0.45	27.8	42.0
CMOS/TD	1.0	1×	1.00	13.0	72.0

Conclusions

The technical barriers necessary to use the tunnel diode in conventional IC processes are being removed. We have shown that molecular beam epitaxy can produce uniform and reproducible III-V layers which set the critical peak current and voltage levels precisely across the wafer. We have taken III-V resonant tunneling diode SPICE models and using TI's CMOS models, provided a first look at the benefits of a Si-based TD combined with CMOS. These results show that tunnel diodes can improve circuit performance *without scaling*. Recently demonstrated MBE-grown Si tunnel diodes provide a path to develop this process.

Acknowledgements

We thank E. Beam, III for his development of the InP epitaxial materials described here. We also gratefully acknowledge the support of AFOSR/DARPA under contract F49620-96-C-0006.

References

(1) R. G. Swartz, "In perspective: the tunnel diode," *IEEE Int. Solid State Circ. Conf.*, pp. 278-280 (1986).

(2) A. Seabaugh, et al. , "Resonant tunneling circuit technology: has it arrived?" *1997 GaAs IC Symp.*, pp. 199-122.

(3) A. C. Seabaugh, B. Brar, T. Broekaert, F. Morris, and G. Frazier, "Resonant Tunneling Analog/Mixed Signal Circuit Technology," unpublished.

(4) T. P. E. Broekaert, et al. "A monolithic 4-bit 2-Gsps resonant tunneling analog-to-digital converter," *IEEE J. Solid State Circ.* **33**, 1342-1349 (1998).

(5) L. D. Nguyen, L. E. Larson, and U. K. Mishra, "Ultra-high-speed modulation-doped field-effect transistors: a tutorial review," *Proc. IEEE*, **80**, 494-518 (1992).

(6) N. Shimizu, et al., "$In_{0.53}Ga_{0.47}As$/AlAs resonant tunneling diodes with switching time of 1.5 ps," *Electronics Lett.*, **31**, 1695-1697 (1995).

(7) P. van der Wagt, A. Seabaugh, and E. Beam, III, "RTD/HFET low standby power memory cell," *IEEE Electron Dev. Lett.* **19**, 7-9 (1998).

(8) K. Morita, et al., "High performance CMOS compatible bistable operation at extremely low supply voltage by a novel Si interband tunnel diode," *56th Dev. Res. Conf.*, pp. 42-43 (1998).

(9) S. L. Rommel, et al., "Room temperature operation of epitaxially grown $Si/Si_{0.5}Ge_{0.5}/Si$ resonant interband tunneling diodes," *Appl. Phys. Lett* in press.

(10) P. van der Wagt, *Proc. IEEE*, in press.

Appendix 18

Optical absorption in alloys of Si, Ge, C, and Sn

B. A. Orner,[a)] D. Hits, and J. Kolodzey
Department of Electrical Engineering, University of Delaware, Newark, Delaware 19716

F. J. Guarin
IBM Microelectronics Division, Hudson Valley Research Park, 1580 Route 52, Hopewell Junction, New York 12533

A. R. Powell
Advanced Technology Materials, 7 Commerce Drive, Danbury, Connecticut 06810

S. S. Iyer
SiBond L. L. C., Hudson Valley Research Park, 1580 Route 52, Hopewell Junction, New York 12533

(Received 13 December 1995; accepted for publication 29 February 1996)

Group IV semiconductor alloy systems offer promise as variable band gap alloys compatible with Si technology. Binary, ternary, and quaternary group IV alloys were grown by molecular beam epitaxy on Si substrates. The fundamental absorption edge was measured by Fourier transform infrared spectroscopy to obtain the optical band gap of the alloys, and the position of the fundamental absorption edge was observed to depend on the experimentally measured alloy composition. Our results indicate a variety of Si-rich group IV alloys with various band gaps are experimentally producible. © *1996 American Institute of Physics.* [S0021-8979(96)05411-4]

INTRODUCTION

The epitaxial growth of random alloys of group IV semiconductors on Si substrates is being investigated by many groups. These materials may permit widespread production of heterostructure devices compatible with existing Si technology.[1,2] Vegard's law predicts that the small atomic size of C compensates for Ge and Sn, which are larger than Si. Thus, alloy systems involving C could potentially be lattice matched to Si at certain compositions.

The $Si_{1-x}Ge_x$ system is miscible for all x, and $Si_{i-x}Ge_x$ alloys are easily formed. The solid solubilities of C in Si and Ge are low (at the melting points, 3×10^{18} cm^{-3} for Si, 1×10^{8} cm^{-3} for Ge).[3] The solid solubility of Sn in Si and Ge is likewise low (5×10^{19} cm^{-3} in Si, 5×10^{20} cm^{-3} in Ge).[4] Thus, most group IV alloys may not be formed at high temperature, but can be produced by low temperature growth techniques such as molecular beam epitaxy (MBE),[5] solid phase epitaxy,[6] and chemical vapor deposition.[7] The resulting materials are metastable and are not limited by the solid solubility of the solutes. The actual upper bound to the amount of C and Sn that can be incorporated into a crystalline group IV alloy is still under investigation, as is the experimental validation of Vegard's law for these systems.

An equally important issue is the behavior of the electronic band structure of these systems under alloying. Theoretical investigations have been accomplished for most of the group IV alloys. Studies involving linear and logarithmic interpolation of the Brillouin zone critical points using 3C–SiC and the cubic phase of the elements as endpoints generally indicate monotonic variations of the band gap and materials which are indirect in k space,[8] though some alloys containing Sn may be direct.[9] $Si_{1-y}C_y$ (Refs. 10–12) and $Si_{1-x-y}Ge_xC_y$ (Ref. 12) in particular have been studied using more complex approaches with counterintuitive results: In spite of the large band gap of diamond, these investigations predict that localized strain will result in a band gap decrease (perhaps with some oscillations) as the carbon content is increased. In fact, some alloys containing over 10% carbon are predicted to be semimetallic.

Experimental measurements of the band gap are plentiful for $Si_{1-x}Ge_x$ (Refs. 13 and 14) but sparse for $Si_{1-x-y}Ge_xC_y$ (Refs. 7, 15 and 16), $Si_{1-y}C_y$ (Ref. 17), and $Ge_{1-y}C_y$ (Refs. 18 and 19) with data being available for only a few compositions and temperatures. Experimental band gap measurements of other group IV alloys have, to our knowledge, not been previously reported. In this article, we report on the band gap of several such alloys.

GROWTH

$Si_{1-x}Ge_x$, $Si_{1-x-y}Ge_xC_y$, $Si_{1-x-y-z}Ge_xSn_yC_z$, and $Si_{1-x-y}Sn_xC_y$ alloys were epitaxially grown by MBE on Si(100) substrates polished on both sides. Separate elemental solid sources were used for each alloy component. The Ge and Sn were thermally heated in crucibles. The Si and C sources were crucible-less, with an electron beam used to heat the Si, and a pyrolytic graphite filament used as the C source. Substrate temperatures were 475 °C for all samples, and growth rates were approximately 1 μm/h.

ANALYSIS

X-ray diffraction (XRD) measurements indicated the presence of strained monocrystalline alloy layers. The presence of strained layers was expected based on the lattice constant predicted by Vegard's law and the metastable strain critical thickness of similarly strained Si-rich $Si_{1-x}Ge_x$/Si films.[1] The alloy lattice constant in the growth direction ranged from 5.431 to 5.461 Å, indicating nearly Si lattice matched alloy layers.

[a)] Electronic mail: orner@ee.udel.edu

The Si, Ge, and Sn contents were measured by Rutherford backscattering spectrometry (RBS), which also indicated the layers were uniform in composition. The C content was inferred from the filament current, calibrated by XRD of binary alloys. The composition of the $Si_{0.92}Ge_{0.08}$ sample was inferred from the lattice constant obtained by XRD. Thickness, which ranged from 0.150 to 0.400 μm, was inferred from growth conditions and was found consistent with RBS measurements.

The optical absorption at photon energies near the band gap was measured at room temperature using Fourier transform infrared spectroscopy (FTIR). Measurements were obtained using a Nicolet 740 optical bench operating in the transmission mode with a quartz beam splitter and PbSe detector.

The transmission data of the samples were ratioed to transmission curves of substrate references to remove substrate absorption and front surface reflection effects. The index of refraction of the layer was expected to be different from that of the substrate. This resulted in an additive offset in the absorption data. This offset was removed by establishing as a zero absorption reference the transmission at photon energies well below the band gap where the absorption curve is flat.

Because the layers were thin, small values of the absorption coefficient (α) were difficult to measure. The noise level of the instrument was reduced by averaging the signal from multiple scans. A ratio of two backgrounds collected with the experimental configuration indicated a noise level within 0.05% of the 100% transmission line. At sample thicknesses of 0.15 μm, this permits measurements of absorption coefficients above 30 cm^{-1}.

Plot of $(\alpha h\nu)^2$ versus photon energy $h\nu$ are often used to determine the optical band gap of a direct semiconductor, and plots of $\sqrt{\alpha}$ vs $h\nu$ are likewise used to find the optical band gap and participating phonon energies of an indirect semiconductor. The region of interest in these plots, however, ordinarily lies well below α=30. Because we could not observe this region, we have adopted the curve fitting approach described below.

The data were fit to theoretical fundamental absorption curves for both a direct band gap semiconductor,

$$\alpha=\begin{cases}\dfrac{K(h\nu_{\mathrm{pt}}-E_G)^{1/2}}{h\nu_{\mathrm{pt}}}, & h\nu_{\mathrm{pt}}\geqslant E_G,\\ 0, \quad h\nu_{\mathrm{pt}}<E_G\end{cases} \tag{1}$$

and an indirect band gap semiconductor,

$$\alpha_a=(h\nu_{\mathrm{pt}}+h\nu_{\mathrm{pn}}-E_G)^2, \tag{2}$$

$$\alpha_e=\exp\left(\frac{h\nu_{\mathrm{pn}}}{kT}\right)(h\nu_{\mathrm{pt}}-h\nu_{\mathrm{pn}}-E_G)^2, \tag{3}$$

$$\alpha=\begin{cases}K(\alpha_e+\alpha_a), \qquad h\nu_{\mathrm{pt}}-h\nu_{\mathrm{pn}}\geqslant E_G\\ K\alpha_e, \quad h\nu_{\mathrm{pt}}+h\nu_{\mathrm{pn}}\geqslant E_G>h\nu_{\mathrm{pt}}-h\nu_{\mathrm{pn},},\\ 0, \quad h\nu_{\mathrm{pt}}+h\nu_{\mathrm{pn}}<E_G\end{cases} \tag{4}$$

where $h\nu_{\mathrm{pt}}$ is the photon energy, $h\nu_{\mathrm{pn}}$ is the energy of the dominant momentum-conserving phonon, E_G is the optical band gap, k is Boltzmann's constant, T is temperature, and K is a prefactor that includes several material and physical constants.

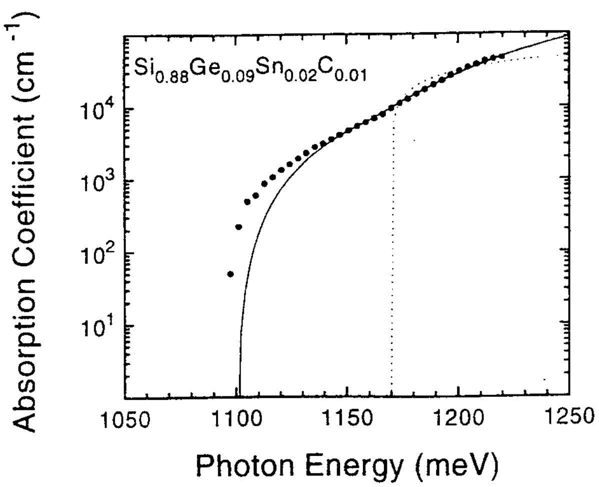


FIG. 1. Experimental fundamental absorption of a $Si_{0.88}Ge_{0.09}Sn_{0.02}C_{0.01}$ thin film (points), and best fit of data using an indirect band gap absorption edge (solid line) and direct band gap absorption edge (dotted line). These fits included absorption data from 1100 to 1230 meV.

These absorption functions were fit to the data using the Levenberg–Marquardt nonlinear curve fitting algorithm as implemented by the ORIGIN[20] software package. Both cases included E_G and K fitting parameters and the indirect case also included $h\nu_{\mathrm{pn}}$.

Figure 1 illustrates the result of direct and indirect band gap curve fits to the data of a $Si_{0.88}Ge_{0.09}Sn_{0.02}C_{0.01}$ sample over a broad energy range; the fit results to the other samples were similar in trends. The shape of the indirect curve matched the data better than the direct curve, though both departed at lower energies. To investigate the cause of this departure, the derivative of α with respect to photon energy was studied. This derivative should be linear for an indirect semiconductor; the sharp increase near 1160 meV shown in Fig. 2 suggests the onset of a second absorption mode, possibly a critical point energy.

In Fig. 3, an indirect curve could be fit well to the data below 1160 meV. (This fit differs from Fig. 1 in that the fit did not include data at energies above 1160 meV). The combination of an indirect curve at low energy and a second indirect curve at higher energies, though, best reflected the experimental data and minimized residual error (χ^2). We considered accounting for the second absorption curve by emission and absorption of an additional momentum conserving phonon in the alloy layer, as well as electron promotion to a second conduction band valley. However, each of the samples measured demonstrated a similar second indirect absorption curves, with onset energies which matched each other within 10 meV.

Therefore we attributed the higher energy indirect absorption curve to residual substrate absorption. This is consistent with the onset, near 1160 meV, of the TO phonon absorption assisted process in Si and the resultant increase in the slope of the absorption edge.[21] The sample spectra had been ratioed to substrates of the same lot number. Although

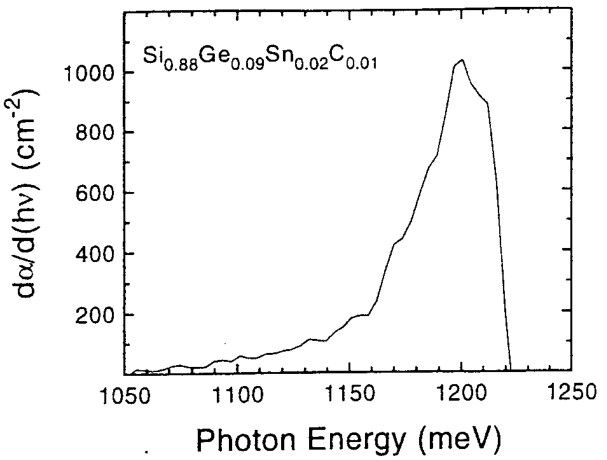


FIG. 2. The derivative of the experimental absorption data of the $Si_{0.88}Ge_{0.09}Sn_{0.02}C_{0.01}$ of Fig. 1. The sharp increase in the slope of the absorption curve near 1160 meV, which is characteristic of Si, was attributed to residual substrate absorption.

TABLE I. Optical band gaps of alloys. Fit results give the momentum conserving phonon energies determined by fitting a theoretical single-phonon fundamental indirect absorption model to experimental data from strained thin films; the stated band gap uncertainties are obtained from the fitting error and assume no uncertainty in the phonon energy. The strain corrected band gap was obtained by deducing the strain in the film from XRD measurements, and adding a correction factor to the fit results. This was compared to the linear estimate of the band gap obtained by interpolation of Brillouin zone critical points (see the text); uncertainties in this column arise from alloy composition uncertainties of ±0.015 for Ge and ±0.005 for C and Sn.

	Fit results		Strain corrected E_G (meV)	Linear estimate E_G (meV)
Composition	$h\upsilon_{ph}$ (meV)	E_G (meV)	E_G (meV)	E_G (meV)
$Si_{0.92}Ge_{0.08}$	43	1099±0.88	1113±0.88	1101±2
$Si_{0.91}Ge_{0.08}C_{0.01}$	54	1119±1.4	1127±1.4	1145±23
$Si_{0.88}Ge_{0.09}Sn_{0.02}C_{0.01}$	57	1145±1.1	1149±1.1	1147±21
$Si_{0.95}Sn_{0.04}C_{0.01}$	48	1145±7.7	1145±7.7	1159±23

we were able to remove most of the substrate absorption, some substrate absorption is likely to remain due to slight thickness variations (Δd_{sub}) between substrates. Since the transmission varies with $\exp(-\alpha d)$, the effect of these thickness variations depended strongly on the substrate absorption coefficient (α_{sub}). Below 1160 meV, $\alpha_{sub} < 0.2$ cm^{-1} and the impact of $\Delta d_{sub} \neq 0$ is negligible even for $\Delta d_{sub} = 100$ μm. Above 1160 meV, α_{sub} rises rapidly (e.g., $\alpha_{sub} = 100$ cm^{-1} at 1300 meV),[21] and significant residual substrate absorption occurs even if $\Delta d_{sub} = 10$ μm. Although an additional indirect absorption curve at high energies (attributable to residual substrate absorption) improved the curve fit as shown by the solid line in Fig. 3, the substrate thickness could not be obtained with sufficient precision (under 1 μm) to allow accurate modeling of the substrate absorption. To avoid introducing errors from this uncertainty, we restricted the curve fits to photon energies below 1160 meV. At those energies, we were confident that any effects of residual substrate absorption were negligible.

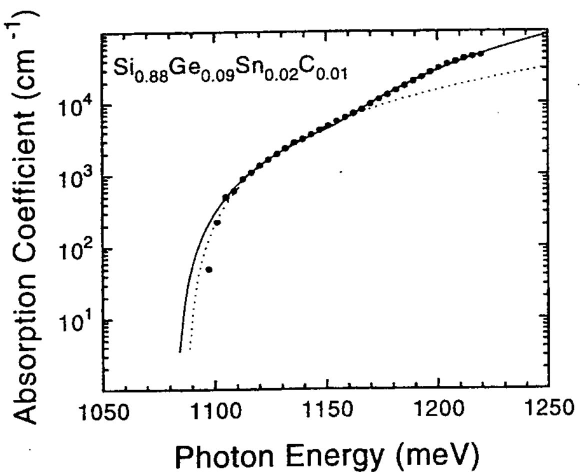


FIG. 3. The best fit of a theoretical indirect curve (dotted line) to the same experimental data as in Fig. 1 (points), with only data below 1160 meV included in the fit. An excellent fit was attained in this region, but the curve departed from the data above 1160 meV as indicated. The addition of another indirect curve representing the residual substrate absorption, beginning at 1160 meV resulted in an excellent fit across a wide range of absorption (solid line).

In general, the fitting algorithm converged upon multiple solutions. Local minima were found with $h\nu_{ph} \approx 5$ meV as well as those with $h\nu_{ph} \approx 50$ meV. Both of these resulted in similar values of χ^2. Previous studies of $Si_{1-x}Ge_x$ alloys[13] indicate that for $0.65 < x < 1.0$ the energy of the dominant phonon is nearly invariant at 48 meV. The phonon energy then drops over a narrow composition range and again remains nearly constant at 23 meV for $0 < x < 0.35$. This trend for $Si_{1-x}Ge_x$ optical phonons has been confirmed in photolumiscence studies.[22] The alloys in our investigation all contain over 65% Si, thus we expect phonon energies near 50 meV. Solutions requiring $h\nu_{ph} \approx 5$ were rejected for their inconsistency with this expectation. After rejecting such solutions, a unique solution remained for each sample.

The E_G and $h\nu_{ph}$ parameters of Eq. (4) exhibit a strong mutual dependence at the experimentally observable values of α. As a result, the fitting uncertainties of these two parameters were similar, though this uncertainty was still under 25 meV for all samples except the $Si_{0.95}Sn_{0.04}C_{0.01}$ sample. If the alloys studied here do in fact behave similar to $Si_{1-x}Ge_x$ alloys and the $h\nu_{ph}$ is considered known, the band gap uncertainties are reduced to the values of Table I.

Figure 4 illustrates the resulting curve of the best fit to the experimental data of each sample. In each case, the indirect model provided the best fit in terms of both form and residual error. The parameters from these curve fits are summarized by Table I. The results for the $Si_{0.92}Ge_{0.08}$ sample agreed well with established experimental values.[21]

Because the alloys were nearly lattice matched to the substrate, strain effects were minimized, but still present. Following St. Amour *et al.*,[16] we obtained the unstrained alloy band gap by computing a strain correction factor. We obtained the alloy lattice constant in the growth direction from XRD, estimated Poisson's ratio (ν) for each alloy by

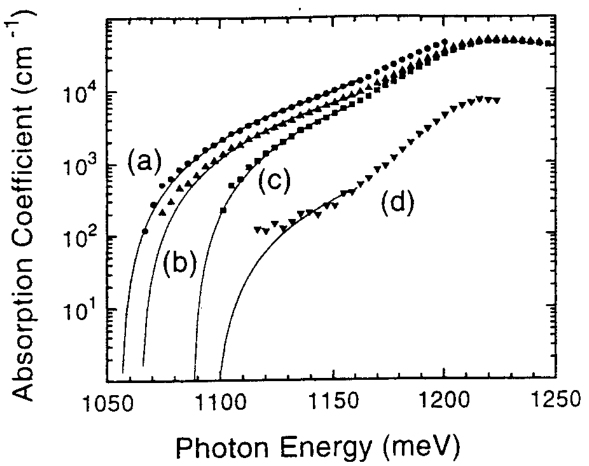


FIG. 4. Experimental fundamental absorption (points) and curve of best fit of thin films of (a) $Si_{0.92}Ge_{0.08}$, (b) $Si_{0.91}Ge_{0.08}C_{0.01}$, (c) $Si_{0.88}Ge_{0.09}Sn_{0.02}C_{0.01}$, and (d) $Si_{0.95}Sn_{0.04}C_{0.01}$. An indirect absorption curve resulted in the best fit in each case.

interpolating between the elemental values (ν=0.10 for diamond, $\nu\approx$0.28 for Ge, Si, and Sn),[21] and computed a biaxial strain in the plane of the heterointerface. Then, the deformation potential was assumed to be similar to that of the $Si_{1-x}Ge_x$ system ($\Delta E_G \approx -11.7$ eV/unit strain),[23] and the resulting correction factors were added to our experimental band gaps (Table I).

In an approach similar to Soref,[8,9] the experimental band gaps of each sample were compared to an estimate based on linear interpolations of Brillouin zone X, L, and Γ critical point conduction band energies,

$$E_G = \min_{i=X,L,\Gamma} [E^i_{\mathrm{Si}}(1-x-y-z)+E^i_{\mathrm{Ge}}x+E^i_{\mathrm{Sn}}y+E^i_{\mathrm{C}}z], \tag{5}$$

where the E^i are the conduction band energy minima at zone position i with respect to the valance band maximum at Γ, and x, y, and z are the atomic fractions of Ge, Sn, and C. Endpoints were based primarily on experimental observations, supplemented by psuedopotential calculations, of the band structure of the diamond phase of the pure elements.[21] This approach predicted conduction band minima at the X points for the composition of all the samples in this study, and band gap energies that were generally in good agreement with our experimental results. The critical point energy near the X point is larger for Sn than for Si, resulting in an increase in the band gap energy when Sn is added to a Si-rich alloy. This was experimentally observed.

CONCLUSION

In conclusion we have grown a variety of random crystalline alloys of Si, Ge, C, and Sn by MBE. The optical band gap was deduced by measuring the fundamental absorption edge of the alloys. These alloys were found to be indirect materials at the compositions studied, and the band gap was found to be nearly linearly dependent on composition. This investigation supports the use of these alloys as variable band gap materials compatible with existing Si technology.

ACKNOWLEDGMENTS

We are grateful to C. P. Swann, Department of Physics and Astronomy, University of Delaware, for his assistance in performing RBS measurements. This work is supported by AASERT Grant No. F49620-92-J-0340, AFOSR Grant No. AFOSR 91-0370, and ONR Grant No. N00014-93-1-0393.

[1] J. C. Bean, Proc. IEEE **80**, 571 (1992).
[2] R. A. Soref, Proc. IEEE **81**, 1687 (1993).
[3] R. I. Scace and G. A. Slack, J. Chem. Phys. **30**, 1551 (1959).
[4] A. G. Milnes, *Deep Impurities in Semiconductors* (Wiley, New York, 1973).
[5] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, Appl. Phys. Lett. **60**, 3033 (1992).
[6] J. W. Strane, H. J. Stein, S. R. Lee, B. L. Doyle, S. T. Picraux, and J. W. Mayer, Appl. Phys. Lett. **63**, 2786 (1993).
[7] P. Boucaud, C. Francis, F. H. Julien, J. M. Lourtoiz, D. Bouchier, S. Bodnar, B. Lambert, and J. Regolini, Appl. Phys. Lett. **64**, 875 (1994).
[8] R. A. Soref, J. Appl. Phys. **70**, 2470 (1991).
[9] R. A. Soref and C. H. Perry, J. Appl. Phys. **69**, 539 (1991).
[10] A. A. Demkov and O. F. Sankey, Phys. Rev. B **48**, 2207 (1993).
[11] J. Gryko and O. F. Sankey, Phys. Rev. B **51**, 7295 (1995).
[12] J. Xie, K. Zhang, and X. Xie, J. Appl. Phys. **77**, 3868 (1995).
[13] R. Braunstein, A. R. Moore, and F. Herman, Phys. Rev. **109**, 695 (1958).
[14] E. R. Johnson and S. M. Christian, Phys. Rev. **95**, 560 (1954).
[15] J. Kolodzey, P. R. Berger, B. A. Orner, D. Hits, F. Chen, A. Khan, X. Shao, M. Waite, S. I. Shah, C. P. Swann, and K. M. Unruh, J. Cryst. Growth **157**, 386 (1995).
[16] A. St. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewalt, Appl. Phys. Lett. **67**, 3915 (1995).
[17] K. Brunner, K. Eberl, and W. Winter, Phys. Rev. Lett. **76**, 303 (1996).
[18] J. Kolodzey, P. A. O'Neil, S. Zhang, B. Orner, K. Roe, K. Unruh, C. Swann, M. Waite, and S. I. Shah, Appl. Phys. Lett. **67**, 1865 (1995).
[19] B. A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R. G. Wilson, P. R. Berger, and J. Kolodzey, J. Electron. Mater. **25**, 297 (1996).
[20] ORIGIN, Microcal Software, Incorporated.
[21] *Landolt-Bornstein: Numerical Data and Functional Relationships in Science and Technology, Group III*, edited by O. Madelung (Springer, Berlin, 1982), Vol. 17a.
[22] J. Weber and M. I. Alonso, Phys. Rev. B **40**, 5683 (1989).
[23] C. G. Van De Walle and R. M. Martin, Phys. Rev. B **34**, 5621 (1986).

J. Appl. Phys., Vol. 79, No. 11, 1 June 1996 Orner *et al.* 8659

Appendix 19

Band gap of Ge rich $Si_{1-x-y}Ge_xC_y$ alloys

B. A. Orner,[a] J. Olowolafe, K. Roe, and J. Kolodzey
Department of Electrical Engineering, University of Delaware, Newark, Delaware 19716

T. Laursen and J. W. Mayer
CSSS, Arizona State University, Tempe, Arizona 85287

J. Spear
Philips Electronic Instruments Company, 125 West Gemini Drive, Suite E-20, Tempe, Arizona 85283

(Received 30 July 1996; accepted for publication 13 August 1996)

$Si_{1-x-y}Ge_xC_y$ films ($x \approx 0.90$, $y \leqslant 0.02$) were grown by molecular beam epitaxy on Si substrates. Infrared optical absorption was used to obtain the band gap energy at room temperature. Biaxial strain obtained from x-ray diffraction measurements verified the presence of nearly relaxed films, and the total and substitutional C contents were obtained from channeling C-resonance backscattering spectrometry. We show by direct measurements that interstitial C had a negligible impact on the band gap, but substitutional C was found to increase the band gap with respect to equivalently strained $Si_{1-x}Ge_x$ alloys. While strain decreases the band gap, the effect of substitutional C on the band gap depends on the Si and Ge fractions. © *1996 American Institute of Physics.* [S0003-6951(96)04343-4]

The energy band gaps of group IV alloys have been studied by photoluminescence (PL),[1,2] absorption,[3–5] spectroscopic ellipsometry,[6,7] and theoretical techniques.[8–10] Several studies suggest an energy band gap decrease as the C content is increased;[6,9] others suggest an increase in the energy band gap.[4,8] Attempts to clarify this behavior have depended on measurements of strained pseudomorphic thin layers with extrapolations to zero strain based on assumption of deformation potentials, which are not well known.

In this letter, we studied relaxed layers with Ge rich compositions. Techniques to probe the indirect optical band gap by absorption are challenged by weaker absorption than a direct band gap material, substrate absorption, and interference fringes within the thin film. We describe an approach to deducing the band gap from optical absorption observations which to some extent transcend these difficulties.

We have produced a series of crystalline $Si_{1-x-y}Ge_xC_y$ alloys by solid source molecular beam epitaxy (MBE) on (100) Si substrates.[4,11] The substrate temperature was 600 °C. Ge was thermally evaporated in a pyrolytic boron nitride crucible at 1260 °C, and Si was thermally evaporated in a pyrolytic graphite (PG) crucible at 1685 °C. the C source consisted of a PG filament heated by a direct current up to 49 A. Growth rates were typically 0.06 μm/h. Thicknesses were measured by a mechanical stylus method and were typically 0.18 μm.

Film compositions (Table I) were measured by Rutherford backscattering spectrometry (RBS). C elastic-resonance backscattering was used to measure the C concentration using a 4.3 MeV He beam. The total measured C concentrations were up to 2 at. %, with an accuracy of ±0.2 at. % C. The extent of substitutional C was studied using ion channeling characterizations. In samples SGC-67, SGC-68, and SGC-70, less than 25% of the C was substitutional, but at least 70% of the total C in SGC-69 was found to be substitutional. Further RBS details will be published elsewhere.[12]

X-ray diffraction (XRD) measurements were performed to determine the strain in the films. The lattice constants of the films in the growth direction ($a_\perp$) were computed from the peak position of the alloy (004) x-ray reflections. The lattice constants in the direction parallel to the substrate surface ($a_\parallel$) were deduced from the (224) and (115) reflections. Assuming Poisson's ratio to be 0.27, we calculated the relaxed cubic lattice constant (a_0) and the biaxial strain for each sample. We observed $a_\perp - a_\parallel < 0.03$ Å for all samples, indicating nearly relaxed films. Relaxed films were expected as the films were well over the 20 Å critical thickness[13] of Ge rich Si alloys.

The optical absorption at photon energies near the band gap was measured at room temperature by Fourier transform infrared (FTIR) spectroscopy. The alloys studied were Ge rich, thus their band gap energies were less than that of Si, and the substrates were transparent in the vicinity of the alloy's fundamental absorption edge. The transmission data of the samples were ratioed to transmission curves of a reference substrate to remove sub-gap substrate absorption errors.

TABLE I. The effect of substitutional and interstitial C on the energy band gap of Ge rich $Si_{1-x-y}Ge_xC_y$ alloys. The Si, Ge, and total C content of each sample was measured by RBS. The strain in each thin film was obtained from XRD, and the experimental band gaps were obtained by curve fitting to the fundamental absorption edge. The band gap of each sample was compared to a hypothetical $Si_{1-x}Ge_x$ alloy with the same Si:Ge ratio and strain (far right column). Over 70% of the C in sample SGC-69 was substitutional; the other samples had less than 20% substitutional C. Thus, substitutional C increased the band gap, and interstitial C had a negligible effect on the band gap.

Sample	Composition	Strain	Band gap (meV)	Equivalently strained $Si_{1-x}Ge_x$ band gap (meV)
SGC-67	$Si_{0.135}Ge_{0.845}C_{0.02}$	1.5×10^{-3}	889±3	896
SGC-68	$Si_{0.11}Ge_{0.88}C_{0.01}$	1.5×10^{-3}	866±5	867
SGC-69	$Si_{0.11}Ge_{0.88}C_{0.01}$	3.0×10^{-3}	895±9	850
SGC-70	$Si_{0.10}Ge_{0.88}C_{0.01}$	1.5×10^{-3}	854±7	856

---

[a] Electronic mail: orner@ee.udel.edu

0003-6951/96/69(17)/2557/3/$10.00

© 1996 American Institute of Physics

A model of thin film absorbance in the presence of interference fringes was applied to deduce an absorption coefficient free of interference fringe effects.[14] The transmittance of a thin absorbing film on a substrate is given by

$$T=\frac{(t_1t_2)^2e^{-\alpha d}}{1+(r_1r_2)^2e^{-2\alpha d}+2r_1r_2e^{-\alpha d}\cos(2\phi-\pi)}, \qquad (1)$$

where r_1, t_1, r_2, and t_2 are the Fresnel reflection and transmission coefficients of the alloy/air and alloy/substrate interfaces, $\phi=2nd(1/\lambda)$, d is the alloy layer thickness, and λ is the vacuum wavelength of the incident photons.

For the purposes of data analysis, α was assumed to be negligible below the energy onset of fundamental absorption. Thus the exponentials of Eq. (1) reduced to unity at sub-gap energies, the amplitude of the fringes was determined solely by the index of refraction (n), and the fringe periodicity (ϕ) was determined by n and d. These parameters were determined by curve fitting the transmittance data in the sub-gap region ($\alpha=0$) to the following expression:

$$T=\frac{(t_1t_2)^2}{1+(r_1r_2)^2+2r_1r_2\cos(2\phi-\pi+\psi)}. \qquad (2)$$

We have introduced a phase shift fitting parameter (ψ), because we observed a phase shift in the interference fringes and were able to more effectively remove the interference from the absorption data by including it. In perfect crystals with real indices of refraction, ψ is expected to be zero. We speculated that the alloy layers were not fully homogeneous, having grain boundaries and voids. Scattering within the alloy layer may be responsible for this phase shift in the interference fringes.

For indirect semiconductors at photon energies below and near the band gap, n is nearly real and constant.[15] We fixed n (hence r_1, t_1, r_2, and t_2) in Eq. (1) at the sub-gap values obtained from the fringes, and Newton's method was applied to solve Eq. (1) for α, given T. This yielded α free from interference fringes.

Because the smaller values of α were obscured by noise, curve fitting was adopted to deduce the band gap from the more strongly absorbing energies. The absorption data were initially fit to

$$\alpha=K(E_G-h\nu_{\mathrm{pt}})^m \qquad (3)$$

with material constant K, band gap E_G, and exponent m taken as adjustable fitting parameters, to determine the mode of the electron transition. Minimum residual error (χ^2) yielded $m\approx 2$for all samples, indicating indirect band gap materials. For sample SGC-69, which had a significant substitutional C content, the residual error was minimized for $1.9<m<2.1$. The fit for the remaining samples converged to $n\approx 2$ less strongly, with comparable values of χ^2 obtained for $1.5<m<2.0$.

Therefore we fixed the exponent at $m=2$ and repeated the curve fit to the following relation:

$$\alpha_a=K(h\nu_{\mathrm{pt}}+h\nu_{\mathrm{pn}}-E_G)^2, \qquad (4)$$

$$\alpha_e=K\left[\exp\left(\frac{h\nu_{\mathrm{pn}}}{kT}\right)(h\nu_{\mathrm{pt}}-h\nu_{\mathrm{pn}}-E_G)^2\right], \qquad (5)$$

$$\alpha=\begin{cases}\alpha_e+\alpha_a & h\nu_{\mathrm{pt}}-h\nu_{\mathrm{pn}}\geqslant E_G,\\ \alpha_a & h\nu_{\mathrm{pt}}+h\nu_{\mathrm{pn}}\geqslant E_G>h\nu_{\mathrm{pt}}-h\nu_{\mathrm{pn}},\\ 0 & h\nu_{\mathrm{pt}}+h\nu_{\mathrm{pn}}<E_G,\end{cases} \qquad (6)$$

where α_a and α_e are the photon absorption contributions from the phonon absorption and phonon emission processes, respectively, $h\nu_{\mathrm{pt}}$ and $h\nu_{\mathrm{pn}}$ are the photon and phonon energies, kT is the thermal energy and K includes several material and physical constants. Adjustable fitting parameters included E_G, K, and $h\nu_{\mathrm{pn}}$.

Absorption data and the best fit for sample SGC-69 are illustrated by Fig. 1. The data points were unweighted, so the fitting algorithm favored data points with large values of α. Figure 2 shows a comparison between SGC-69 and SGC-68, a sample with predominantly interstitial C. A sub-gap absorption tail is evident in SGC-68, with the theoretical ab-

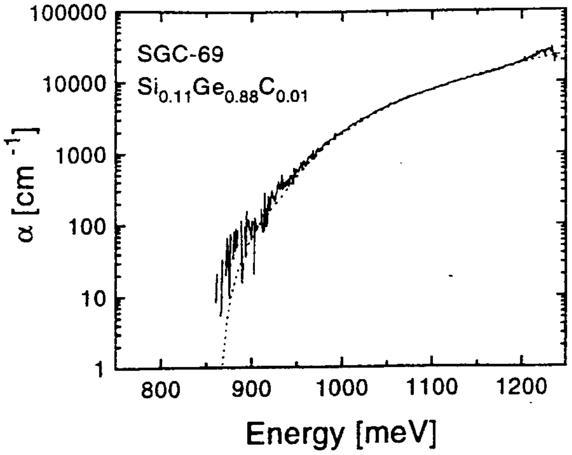


FIG. 1. Optical absorption of SGC-69. Experimental data are indicated by a solid line and the best fit of a theoretical indirect absorption curve is indicated by a dashed line. An excellent fit was obtained, and a band gap of 895±9 meV was deduced from this curve fit.

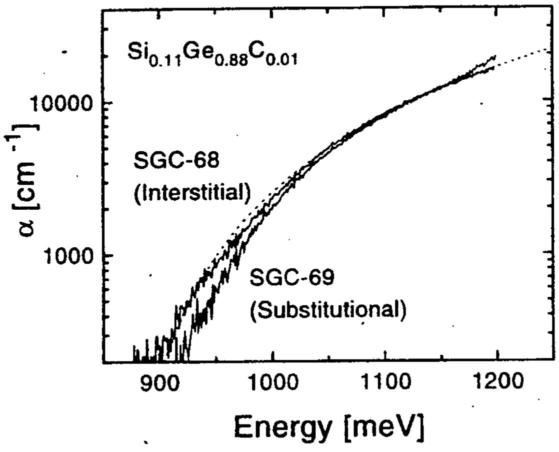


FIG. 2. Optical absorption of SGC-68. This sample has the same composition as the sample in Fig. 1, however the C is primarily interstitial. The solid line indicates experimental data and the dashed line indicates the best fit of an indirect absorption curve. The band gap of this sample is lower, and an absorption tail near $\alpha=100\ \mathrm{cm}^{-1}$ was evident by the departure of the experimental data from the theoretical absorption curve at low energies. The absorption curve of SGC-69 was included for comparison.

sorption curve departing from the experimental data at low energies. Results for SGC-67 and SGC-70 were similar. The absorption tail was not present in SGC-69 (Fig. 1), which has the same composition as SGC-68 but a higher substitutional C content. Fitting results indicated $h\nu_{\mathrm{pn}} \approx 25$ meV for all samples.

During the growth of SGC-69 the C filament was off, thus we conclude that the source of C was the PG crucible used to contain the Si. We speculate that the co-evaporation of Si and C from the crucible favored substitutional C. The C filament was heated to approximately 2000 °C during the growth of the other samples. Further study is necessary to determine the mechanism by which the presence of this C molecular beam caused a decrease in substitutional C.

For Si rich, strained $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$ thin films on Si substrates, C can be viewed as altering the band structure through strain reduction.[1] But the Ge rich alloys studied here are nearly relaxed. Thus bulk strain reduction is not predominantly responsible for the shift in the absorption edge we observed.

The band gap energy of each sample was compared to a hypothetical $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ alloy with the same Si:Ge ratio and the same strain. The band gap was computed as $E_G(L) = 0.7596 + 1.0860(1-x) + 0.3306(1-x)^2$.[15] The small residual strain measured by XRD for each $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$ sample was applied to an established deformation potential relation for the strained $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ system[16] to obtain the hypothetical $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ band gap with the same strain as our $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$ alloys (Table I). With this approach, it is not necessary to assume a deformation potential for $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$, which is not yet known. The samples with predominately interstitial C had band gaps similar to the equivalently strained $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ alloy. In SGC-69, the substitutional C was found to increase the energy band gap by 45 meV/%C, which we attribute to the effect of alloying.

We have grown a series of $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$ thin films ($x \approx 0.90$, $y \leqslant 0.02$) by MBE and conducted an experimental study of their energy band gaps. The location of the C in the crystal was found to impact the energy band gap of the material, which supports the assertion that C may be used to tune the band gap. Substitutional C was found to increase the band gap at this composition. Other workers have found C decreases the band gap with respect to a $\mathrm{Si}_{1-x}\mathrm{Ge}_x$ alloy. However, other work has been limited to Si rich alloys, and the band gap may in fact decrease with increasing substitutional C at those compositions. Thus, this work is not necessarily inconsistent with previous results.

The location of C within a Ge rich $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$ lattice is evidently dependent on the growth conditions. The ability to manipulate the growth conditions to increase the substitutional C incorporation will allow the $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$ band gap and lattice constant to be engineered over a wider composition range.

We are grateful to acknowledge the support of K. M. Unruh, S. S. Iyer, and R. Soref for useful discussions. This work is supported by AASERT Grant No. F49620-92-J-0340, AFOSR Grant No. AFOSR 91-0370, and ONR Grant No. N00014-93-1-0393.

[1] A. St. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewalt, Appl. Phys. Lett. **67**, 3915 (1995).

[2] K. Brunner, K. Eberl, and W. Winter, Phys. Rev. Lett. **76**, 303 (1996).

[3] R. Braunstein, A. R. Moore, and F. Herman, Phys. Rev. **109**, 695 (1958).

[4] B. A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R. G. Wilson, P. R. Berger, and J. Kolodzey, J. Electron. Mater. **25**, 297 (1996).

[5] B. A. Orner, D. Hits, J. Kolodzey, F. J. Guarin, A. R. Powell, and S. S. Iyer, J. Appl. Phys. **79**, 8656 (1996).

[6] W. Kissinger, H. Osten, M. Weidner, and M. Eichler, J. Appl. Phys. **79**, 3016 (1996).

[7] S. Zollner, J. Appl. Phys. **78**, 5209 (1995).

[8] R. A. Soref, J. Appl. Phys. **70**, 2470 (1991).

[9] A. A. Demkov and O. F. Sankey, Phys. Rev. B **48**, 2207 (1993).

[10] J. Xie, K. Zhang, and X. Xie, J. Appl. Phys. **77**, 3868 (1995).

[11] J. Kolodzey, P. A. O'Neil, S. Zhang, B. Orner, K. Roe, K. Unruh, C. Swann, M. Waite, and S. I. Shah, Appl. Phys. Lett. **67**, 1865 (1995).

[12] D. Hits, B. A. Orner, J. Olowolafe, J. Kolodzey, T. Laursen, and J. Mayer (unpublished).

[13] J. C. Bean, Proc. IEEE **80**, 571 (1992).

[14] W. H. Brattain and H. B. Briggs, Phys. Rev. **75**, 1705 (1949).

[15] *Landolt-Bornstein: Numerical Data and Functional Relationships in Science and Technology, Group III*, edited by O. Madelung (Springer, Berlin, 1982), Vol. 17a.

[16] C. G. Van De Walle and R. M. Martin, Phys. Rev. B **34**, 5621 (1986).

Current Transport Characteristics of SiGeC/Si Heterojunction Diode

F. Chen, B. A. Orner, D. Guerin, A. Khan, P. R. Berger, *Member, IEEE*,
S. Ismat Shah, and J. Kolodzey, *Senior Member, IEEE*

Abstract— **The characteristics of heterojunction diodes fabricated from p-type epitaxial $Si_{0.07}Ge_{0.91}C_{0.02}$ alloy grown by molecular beam epitaxy on n-type Si (100) have been examined by using current–voltage, capacitance–voltage, and Hall effect measurements. The SiGeC/Si heterojunction diode shows good rectification with nearly ideal forward bias behavior and low reverse leakage currents compared to Ge/Si heterojunction diodes. The temperature dependence of the current-voltage behavior indicates that the principle conduction mechanism is by electron injection over a barrier. Reverse breakdown occurs by the avalanche mechanism.**

I. INTRODUCTION

RECENTLY there has been a growing interest in heterojunction (HJ) devices compatible with silicon. Doped layers of $Si_{1-x}Ge_x$ alloys have improved the performance of bipolar transistors [1] and have produced light emitting diodes [2]. SiGe alloys, however, are lattice-mismatched to Si substrates. The addition of C to SiGe helps with strain compensation, and affects the interface, bandgap energy and the energy band offsets [3]. Although the structural and optical properties of $Si_{1-x-y}Ge_xC_y$ alloys have been reported [4], [5], the electrical properties are not yet well understood.

This paper examines the electrical properties versus temperature of $Si_{1-x-y}Ge_xC_y/Si$ p^+-n heterojunction diodes fabricated from mesa-etched epitaxial layers grown on n-type Si-substrates by molecular beam epitaxy (MBE). We focus on the behavior of Ge-rich layers of composition $Si_{0.07}Ge_{0.91}C_{0.02}$ heavily doped with boron to produce p-type conductivity. We have shown that the valence band offset (ΔE_v) between bulk $Si_{0.07}Ge_{0.91}C_{0.02}$ and Si is about 0.60 eV [6], and this is large enough to impede the injection of holes from SiGeC into Si.

II. EXPERIMENTAL DETAIL

The SiGeC layers were grown in an EPI 620 MBE system on 75-mm diameter (100) oriented Si substrates at a temperature of 550 °C as previously described [4]. Solid elemental sources were used, and the boron cell temperature was 1550 °C for the doping concentration used here. The relatively low substrate growth temperature minimizes the interdiffusion between the layer and substrate. The heterojunction, therefore, is expected to be abrupt. The doped SiGeC layer reported here was 0.18-μm thick and the doped Ge layer was 0.5 μm. Electron diffraction showed the SiGeC layer to be single crystal. From Hall effect measurements, the hole concentrations were 1×10^{19} cm^{-3} in the SiGeC and 5×10^{18} cm^{-3} in the Ge. The hole mobility of $Si_{0.07}Ge_{0.91}C_{0.02}$ at room temperature (90 cm^2/V·s) was somewhat lower than in pure Ge perhaps due to additional alloy scattering, but it was 1.5 times higher than the mobility in pure Si. From conductivity measurements, the n-type Si substrate doping was about 2.5×10^{15} cm^{-3}. Optical absorption measurements of the doped $Si_{0.07}Ge_{0.91}C_{0.02}$ indicated a bandgap of 0.80 eV, but this value may be larger than the effective bandgap because of impurity band absorption. Standard optical lithography and wet chemical etching were used to form the diodes shown schematically in Fig. 1. The SiGeC layers were mesa etched in a $H_3PO_4 : H_2O_2 : H_2O$ (1 : 6 : 3) solution for 1.5 min to a depth of 0.18 μm. Electron beam evaporated Ti/Au metal layers of thickness 300 Å/1500 Å were used for both the p- and n-contacts without annealing. The contacts were patterned by the lift-off method. The mesa area was 2.2×10^{-3} cm^{-2}. The Ti/Au were measured to have ohmic, linear current voltage characteristics over the full measurement range.

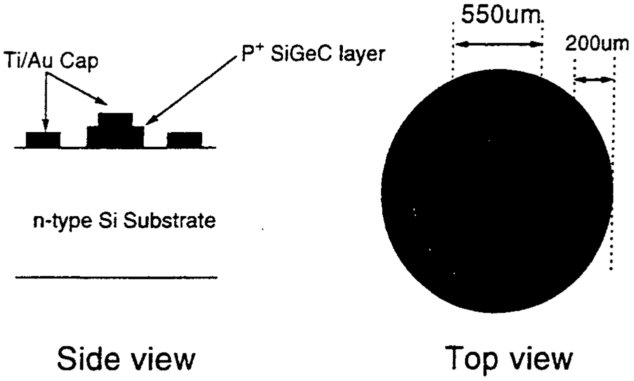


Fig. 1. Schematic cross section and top view of the mesa etched diodes fabricated in 0.18-μm thick p-SiGeC layer on n-Si (100) substrate.

III. RESULTS AND DISCUSSIONS

To provide information on the transport mechanisms, the current-voltage (*I–V*) characteristics were measured versus temperature. Fig. 2 shows the *I–V* characteristics for three

---

Manuscript received June 19, 1996; revised August 6, 1996. This work was supported by the AFOSR under Grant F49620-95-1-0135, the ARO under Grant DAAH04-95-1-0625, the ONR under Grant N00014-93-1-0393, the DARPA contract F49620-96-C-006, and by W. L. Gore & Associates.

F. Chen, B. A. Orner, A. Khan, P. R. Berger, and J. Kolodzey are with the Department of Electrical Engineering, University of Delaware, Newark, DE 19716 USA.

D. Guerin is with the Department of Physics and Astronomy, University of Delaware, Newark, DE 19716 USA.

S. I. Shah is with the Dupont Experimental Station, Wilmington, DE 19898 USA.

Publisher Item Identifier S 0741-3106(96)08869-6.

0741–3106/96$05.00 © 1996 IEEE

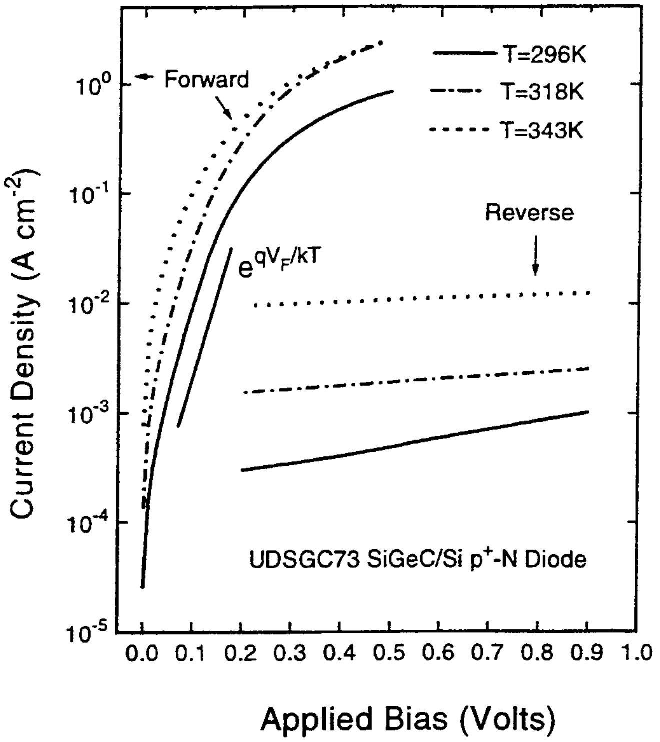


Fig. 2. *I–V* characteristics of SiGeC/Si p$^+$-n heterojunction diode. Diodes are strongly rectifying with nearly ideal behavior at room temperature. In forward bias, ideality factor near unity implies injection over a barrier.

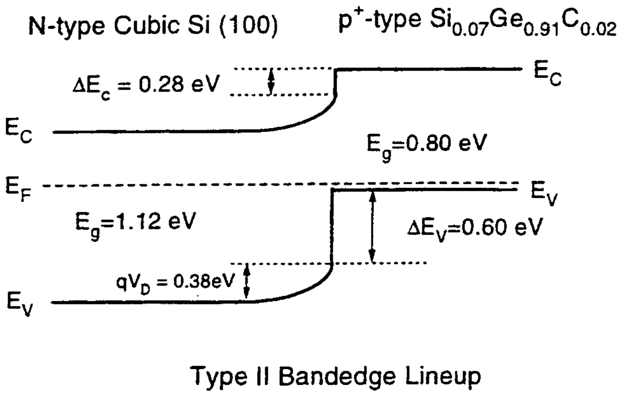


Fig. 3. Energy-band diagram of $Si_{0.07}Ge_{0.91}C_{0.02}$/Si heterojunction. Most of band bending occurs in Si substrate since the SiGeC is more heavily doped. The diagram is calculated using $\Delta E_V = 0.60$ eV and $E_g = 0.80$ eV. Energy band alignment is staggered, type II.

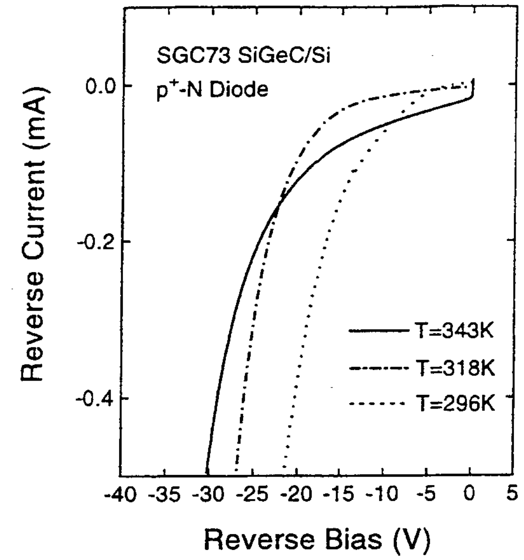


Fig. 4. Temperature dependence of SiGeC/Si heterojunction breakdown voltage, implying an avalanche mechanism.

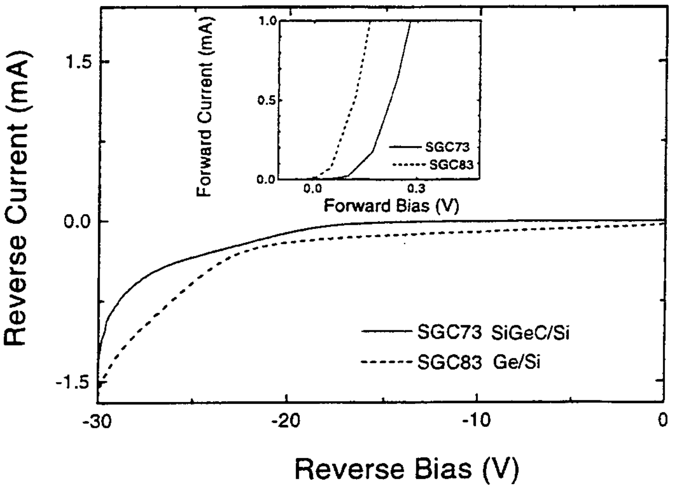


Fig. 5. *I–V* characteristics for SiGeC/Si and Ge/Si diodes. The SiGeC/Si diode has *I–V* characteristics with smaller leakage current, higher breakdown voltage, and higher turn-on voltage.

temperatures from 296 K to 343 K. At low bias at room temperature, the characteristics were nearly ideal with $\eta = 1.1$, where $I = I_0 \exp(qV/\eta kT)$. The deviation from linearity at high forward current may be due to series resistance and high level injection. The forward turn-on voltage is relatively low compared to a pure Si homojunction. Low forward turn-on voltages in heterojunctions are associated with low bandgaps and low values for one of the band offset energies [7], [8]. The forward current increases exponentially with temperature implying thermally activated behavior. The SiGeC/Si diodes exhibit strong rectifying behavior. At biases more positive than -3 V, the reverse current magnitude increases with temperature. At biases more negative than -3 V, the effect of the breakdown voltage, which increases in magnitude with temperature, causes the temperature curves to cross. A plot of $I(dV/dI)$ versus I at $T = 296$ K gives the ideality factor $\eta = 1.1$ from the intercept, and the series resistance $r_s = 138\ \Omega$ from the slope. From the value of η, we infer that the principal forward current mechanism is due to carrier injection over a barrier. Assuming an exponential dependence of saturation current of I_o on temperature, where $I_o \sim \exp(-\Delta E_{AF}/kT)$, the activation energy ΔE_{AF} for the current was obtained from an Arrhenius plot of current versus temperature from 296 K to 343 K at different forward biases. Extrapolation to zero bias yields an activation energy of 0.38 eV. We note that this value is approximately $E_g/2$ using the value $E_g = 0.8$ eV measured from optical absorption, and may indicate the activation energy for electron generation/recombination current in the diode.

This value is also consistent with capacitance–voltage (*CV*) measurements of $1/C^2$ versus V which extrapolate to yield the built-in voltage V_{bi} of 0.38 V. The linearity of the $1/C^2$-V characteristics implied a uniformly-doped abrupt junction. An energy-band diagram of a $Si_{0.07}Ge_{0.91}C_{0.02}$/Si heterojunction, based on the experimental valence-band offset and bandgap data, is shown in Fig. 3. The conduction-band discontinuity $\Delta E_C = 0.28$ eV and the valence-band discontinuity $\Delta E_V = 0.60$ eV. Since ΔE_V is large enough to impede the injection of holes from SiGeC to Si, and is larger than ΔE_C, therefore, we expect that the transport current is mostly due to the injection of electrons from the n-Si into the p-SiGeC.

The reverse current characteristics versus temperature near breakdown are shown in Fig. 4. The breakdown voltage near -20 V corresponds to a critical field $\epsilon_{\text{cri}} = 2.2 \times 10^5$ V/cm which is between the values for Si (3×10^5 V/cm) and Ge (10^5 V/cm). At $T = 296$ K, the prebreakdown current obeys the $I_R \sim (V)^m$ where $m \approx 2.5$ which we attribute to tunneling and surface leakage. At higher temperature, however, the thermal current with $m \approx 1$ dominates. The voltage at which reverse breakdown occurred increased with temperature suggesting an avalanche mechanism.

We compared our SiGeC/Si HJ diode to Ge/Si and SiGe/Si HJ diodes. Fig. 5 shows the *I–V* characteristics for both SiGeC/Si and Ge/Si HJ diodes with the same device area and similar doping level. The SiGeC/Si diode yielded *I–V* characteristics with smaller leakage current, higher breakdown voltage and higher turn-on voltage. For comparison, Kamins [8] *et al.* reported SiGe/Si diodes with reverse leakage current density $J_R = 1 \times 10^{-7}$ Acm^{-2} at bias $= -5$ V compared to 10^{-3} Acm^{-2} for our diodes. The SiGe diodes have the forward current density $J_F = 5 \times 10^{-6}$ Acm^{-2} at bias $= 0.2$ V compared to 10^{-1} Acm^{-2} for our SiGeC diodes. The Kamins diodes, however, were Si-rich (22% Ge) and our diodes are Ge-rich (91% Ge).

IV. CONCLUSIONS

The characteristics of $Si_{1-x-y}Ge_xCy$/Si diodes and their temperature dependence were investigated. Current–voltage characteristics were nearly ideal at low forward bias at room temperature. The temperature behavior indicated a barrier height for carrier injection of 0.38 V. Due to expected large valence band offsets, we conclude that the conduction mechanism is the injection of electrons from the wider gap n-Si to the p$^+$-SiGeC layer. These results indicate that the SiGeC/Si heterostructure has good electrical properties and is promising for Si-based device applications.

REFERENCES

[1] A. Gruhle, H. Kibbel, U. Konig, U. Erben, and E. Kasper, "MBE-grown Si/SiGe HBT's with high β, f_T, and $f_{\max}$," *IEEE Electron Device Lett.*, vol. 13, pp. 206–208, 1992.

[2] J. Sturm, A. A. St. Amour, Q. Mi, L. C. Lenchyshyn, and M. L. W. Thewalt, "High temperature (77–300 K) photo- and electroluminescence in Si/$Si_{1-x}Ge_x$ heterostructures," *Jpn. J. App. Phys.*, vol. 33, pp. 2329–2334, 1994.

[3] A. R. Powell, K. Eberl, B. A. Ek, S. S. Iyer, "$Si_{1-x-y}Ge_xC_y$ growth and properties of the ternary system," *J. Cryst. Growth*, vol. 127, pp. 425–429, 1993.

[4] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, "Growth of germanium-carbon alloys on silicon substrate by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 67, pp. 1865–1867, 1995.

[5] B. A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R. G. Wilson, P. R. Berger, and J. Kolodzey, "Optical properties of $Ge_{1-y}C_y$ alloys," *J. Electron. Mater.*, vol. 25, pp. 297–300, 1996.

[6] F. Chen, M. M. Waite, S. Ismat Shah, B. A. Orner, S. S. Iyer, and J. Kolodzey, "Measurements of the energy band offsets of $Si_{1-x}Ge_x$/Si and $Ge_{1-y}C_y$/Ge heterojunctions," *Appl. Surf. Sci.*, vol. 104–105, pp. 615–620, 1996.

[7] T. I. Kamins, K. Nauka, R. D. Jacowitz, J. L. Hoyt, D. B. Noble, J. F. Gibbons, "Electrical and structure properties of diodes fabricated in thick, selectively deposited Si/$Si_{1-x}Ge_x$ epitaxial layers," *IEEE Electron Device Lett.*, vol. 13, pp. 177–179, 1992.

[8] J. D. Hwang, Y. K. Fang, K. H. Chen, and D. N. Yaung, "A novel β-SiC/Si heterojunction backward diode," *IEEE Electron Device Lett.*, vol. 16, pp. 193–195, 1995.

Dielectric response of thick low dislocation-density Ge epilayers grown on (001) Si

Kelly E. Junge, Rüdiger Lange, Jennifer M. Dolan, and Stefan Zollner[a)]
Ames Laboratory and Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

M. Dashiell, B. A. Orner, and J. Kolodzey
Department of Electrical and Computer Engineering, 140 Evans Hall, University of Delaware, Newark, Delaware 19716

(Received 16 September 1996; accepted for publication 24 October 1996)

Spectroscopic ellipsometry was used to measure the dielectric functions of epitaxial and bulk Ge at photon energies from 1.5 to 5.2 eV. The epitaxial Ge was grown at 400 °C by molecular beam epitaxy on (001) Si substrates. The optical response and the interband critical-point parameters of Ge on Si were found to be indistinguishable from that of bulk single crystal Ge, indicating high optical quality. Dislocation density measurements using an iodine etch verified low surface defect densities. We conclude that epitaxial Ge grown on Si at relatively low temperatures is suitable for optical device applications. © *1996 American Institute of Physics.* [S0003-6951(96)04352-5]

Because of the large lattice mismatch (~4%) between Si and Ge and their different thermal expansion coefficients, it is difficult to produce high-quality heteroepitaxial Ge films on Si substrates. On the other hand, the growth of low defect-density Ge[1,2] and $Si_{1-x-y}Ge_xC_y$ alloys[3–5] on Si (001) is of great technological interest to modify transport[3,4,6] and carrier confinement in Si-based devices and to integrate Ge-based optoelectronics (e.g., infrared detectors) with well-developed Si integrated circuit technology. Also, because of the small lattice mismatch of GaAs and Ge, it becomes feasible to integrate III/V devices with Si integrated circuit technology by growing a Ge virtual substrate on Si, followed by GaAs/Ge epilayers.[7]

The initial growth of Ge on Si (001) can be described[8] as Stranski-Krastanow: The first three to six monolayers grow layer-by-layer. After about 10 to 15 Å , islands start to form. Island formation can be suppressed by a surfactant (As), extending the layer-by-layer growth.[8] Because of the local elastic deformation of near-surface layers in the substrate, the onset of dislocations is delayed[8] until the islands have grown to far in excess (500 Å) of the equilibrium critical thickness (10 Å).

This work deals with much thicker layers, where the strain is relieved by misfit dislocations. These dislocations can be studied using plan-view transmission electron microscopy (TEM).[2,4,9] They usually climb to the surface where they deteriorate the transport properties of active device layers (such as a heterojunction bipolar transistor[4]). Using thick compositionally graded layers,[10] it is possible to grow relaxed $Si_{1-x}Ge_x$ layers ($x\sim0.3$) with low threading dislocation densities.

Malta *et al.*[2] have shown that dislocations in Ge on Si can be confined to the epilayer/substrate interface (extending up to 0.7 μm from the interface) by growing with substrate temperatures near the melting point of Ge (937 °C). For samples with a thickness of 2.5 μm, the residual strain was $e\sim2.5\times10^{-3}$ and the etch pit density (EPD), a measure for the dislocation density at the surface, was about 2×10^5 cm^{-2}. Apparently, interfacial Ge melts and subsequently alloys with the Si substrate. Growth at intermediate temperatures (700 °C) does not confine the dislocations, resulting in a higher EPD.

The purpose of this work is to show that slow growth at low temperatures can yield thick Ge films on Si with low surface dislocation densities. This leads to the surprising result that the dielectric function (DF) of Ge on Si is indistinguishable from that of a bulk Ge sample. (The correlation between the DF and of dislocations in group-IV alloys was discussed by Lange *et al.*[11] Thick relaxed $Si_{1-x-y}Ge_xCy$ layers with many dislocations have very broad and weak E_1 peaks, whereas pseudomorphic layers with a low dislocation density have strong and narrow E_1 peaks.)

The Ge layers used in this study were grown[12] by molecular beam epitaxy (MBE). The Ge molecular beam was produced by thermal evaporation from a solid source of zone-refined polycrystalline Ge in a pyrolytic boron nitride crucible. To minimize contamination from the crucible, the cell temperature was kept below 1380 °C. At the cell temperature of 1325 °C used in this experiment, the Ge growth rate was 0.11 μm/h (0.3 Å/s).

Substrates were (001) oriented, 75 mm diameter silicon wafers prepared by degreasing, oxidizing in a solution of H_2O:H_2O_2:HCl (5:3:3), and dipping in HF:H_2O (1:10).[12] The substrates were desorbed at 250 °C in the MBE chamber just prior to growth. The Ge layers were grown at a substrate temperature of 400 °C; they were between 0.3 and 1.1 μm thick (measured by a Dektak) and appeared mirror smooth after growth. Since the mobility of dislocations is limited at lower temperatures[9] and the thermal expansion coefficient of Ge (6×10^{-6}) is about a factor of two larger than that of Si, it is not surprising that growth at low temperatures resulted in high quality layers.

Well resolved reciprocal lattice rods of the substrate were observed by *in situ* reflection high energy electron diffraction (RHEED). After about 10 Å of growth, the lattice rods were less well resolved, but distinct rods still remained. The RHEED pattern gradually improved as the epilayer became thicker. For sample SGC99, a 0.75 μm thick layer of

---

[a)] Electronic mail: zollner@iastate.edu

© 1996 American Institute of Physics

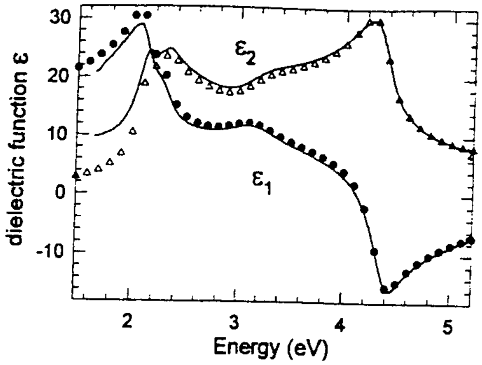


FIG. 1. Lines: Real (ϵ_1) and imaginary (ϵ_2) part of the dielectric function of SGC99 (0.75 μm Ge on Si), corrected for a 10 Å native oxide layer. The data of Ref. 13 for bulk Ge ⟨111⟩ are shown for comparison (●),(△).

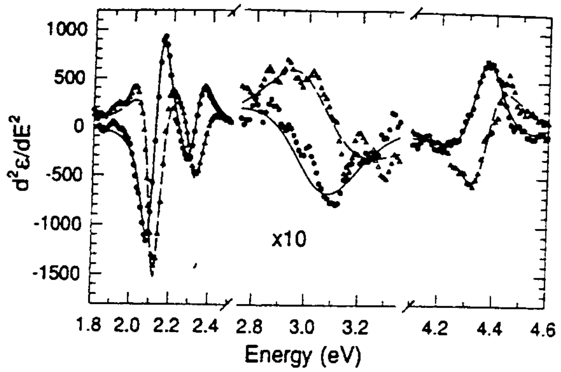


FIG. 2. Numerically calculated second derivatives of ϵ_1 (●) and ϵ_2 (△) for Ge on Si. The lines give the best fit to Eq. (1) with the parameters in Table I. The E_0' region (2.75–3.35 eV) was multiplied by 10 to make it visible on this scale.

Ge on Si grown at 400 °C, the RHEED pattern at the completion of growth was similar in features and intensity to that of commercially available Ge substrates. The RHEED suggested that island formation was partially suppressed at low growth temperatures, and that as growth proceeds islands may coalesce to form single crystal Ge with few defects. We speculate that the low growth rates employed here encourage the formation of reduced defect single crystal Ge over multicrystalline Ge, but further study will be necessary to confirm this.

To find the surface dislocation densities, we used an iodine etch[2] HF:HNO$_3$:CH$_3$COOH:I (20 ml:40 ml:44 ml:120 mg) for 1 s to measure the etch pit density (EPD) of the Ge layers. For SGC99, it appeared constant and uniform across the entire area and was consistent between samples and etch times. The average EPD was 4×10^4 cm^{-2}, a factor of five lower than the results of Malta *et al.*[2] The EPDs of thinner layers (<0.3 μm) and those of samples grown at higher temperatures (>500 °C) could not be determined, since the EPD was not uniform or the complete Ge layer was removed by the etch. The EPD of bulk Ge was less than 10^4 cm^{-2}, consistent with data supplied by Eagle Picher. The pit shapes for SGC99 and bulk Ge differed. For the bulk Ge, most pits were circular, about 1 μm in diameter. For SGC99, the pits were squares, approximately $1-3$ μm on each side.

After growth, the dielectric functions (DFs) ϵ in the 1.5 to 5.5 eV photon-energy range were measured *ex situ* with a spectroscopic ellipsometer.[13] The spectra were corrected for a native oxide layer. The thickness of the oxide was determined by matching ϵ_2 at its peak near 4.2 eV with the data of Ref. 13. The lines in Fig. 1 show the real (ϵ_1) and imaginary (ϵ_2) parts of ϵ for sample SGC99, assuming an oxide thickness of 10 Å. Other Ge epilayers grown on Si at the same temperature (not shown in the figure) had similar ϵ. For comparison, we also measured ϵ for a commercial bulk Ge ⟨001⟩ sample (Eagle Picher). The DF of SGC99 and that of the bulk sample were indistinguishable, except below 1.8 eV, where the accuracy of our instrument decreases. In Fig. 1 we also show the data of Ref. 13 (●), (△) for bulk ⟨111⟩ Ge. The agreement is good, except for ϵ_2 in the range below 2 eV. (Similar discrepancies were found in Ref. 16.) The DF of SGC99 resembles that of bulk Ge much more than that of thin Ge films enclosed between Si barriers.[14,15]

The spectra show a double-peak structure above 2 eV (E_1, $E_1+\Delta_1$), a shoulder near 3 eV (E_0'), and a third peak near 4.2 eV (E_2). These peaks are interband critical points (CPs) arising from direct band-to-band transitions at various regions in the Brillouin zone.[17] For a further analysis of these CPs, we calculate numerically the second derivative of ϵ with respect to photon energy (shown by the symbols in Fig. 2) and perform a line shape analysis. Following Viña *et al.*,[17] we describe the CPs using a mixture of a 2D minimum and a saddle point represented by

$$\epsilon(\omega)=C-A\ \ln(\hbar\omega-E_g-i\Gamma)\exp(i\phi), \qquad (1)$$

where $\hbar\omega$ is the photon energy, E_g the energy of the CP, Γ its broadening, A its amplitude (oscillator strength), and ϕ the phase angle describing the amount of mixing. The parameters obtained from the line shape analysis are given in Table I in comparison with parameters of bulk samples from Viña and co-workers.[17] First, we note that our bulk parameters are, within the error bars, identical to those of Ref. 17 with one exception: Viña and co-workers used a fixed spin-orbit splitting $\Delta_1=187$ meV determined from low-temperature measurements. In our analysis, we treated Δ_1 as

TABLE I. Critical point (CP) parameters for bulk Ge and Ge on Si: amplitude (A), energy (E), broadening (Γ), and excitonic phase (Φ) [see Eq. (1)].

	A (1)	E (eV)	Γ (eV)	Φ (deg)
		Bulk Ge (this work)		
E_1	5.5(3)	2.114(2)	0.058(2)	86(4)
$E_1+\Delta_1$	4.1(6)	2.314(2)	0.076(6)	same
E_0'	3.2(6)	3.05(2)	0.20(2)	−29(12)
E_2	8(1)	4.37(1)	0.107(1)	−193(11)
		Bulk Ge (from Ref. 17)		
E_1		2.111(3)	0.06(1)	71(4)
$E_1+\Delta_1$		2.298(3)	0.07(2)	same
E_0'		3.11		
E_2		4.368(4)	0.109(9)	
		Ge on Si (SGC99, this work)		
E_1	6.2(4)	2.116(2)	0.063(2)	84(4)
$E_1+\Delta_1$	3.7(7)	2.322(2)	0.076(6)	same
E_0'	3.3(5)	3.05(2)	0.21(2)	−29(9)
E_2	8(1)	4.37(1)	0.109(6)	−196(6)

a free parameter (since it is a measure for the strain in the sample) and found $\Delta_1 = 200$ meV for bulk Ge.

The CP parameters for sample SGC99 are similar to those of bulk Ge. Most importantly, the broadenings, related to defects, are essentially the same. Therefore, the scattering of electrons and holes in SGC99 was mostly due to intrinsic mechanisms such as electron-phonon interactions, not to sample imperfections such as dislocations, grain boundaries, impurities, etc. The spin-orbit splitting parameter for SGC99 was $\Delta_1 = 206$ meV, about 3% larger than in bulk Ge. Using the small-shear approximation described in Ref. 11, we found upper bounds for the hydrostatic and (001) shear strains (e_H and e_S) in SGC99. Since E_1 is the same for bulk Ge and SGC99, we conclude that the hydrostatic and (001) shear shifts for E_1 (ΔE_H and ΔE_S) are approximately equal. Since (the apparent splitting) Δ_1 changes by no more than 6 meV, ΔE_H and ΔE_S are about 3 meV each. We conclude that $|e_H| < 0.03\%$ and $|e_S| < 0.1\%$. Since $\Delta E_H \propto e_H$, whereas $\Delta E_S \propto e_S^2$ our estimate for e_S is less stringent than that for e_H. Using x-ray diffraction, the in-plane strain perpendicular to the growth axis ($e_\perp = e_H - e_S$) was determined for similar samples[5] to be below 0.03%, about three times smaller than the upper limit found here. Although our accuracy is limited, we find less than 3% of the strain expected for a pseudomorphic layer (equal to the lattice mismatch of 0.04). The accuracy of our strain analysis could be improved by measuring ϵ below 100 K (where the broadenings are smaller leading to more accurate CP energies).

In conclusion, we have found that the optical constants (refractive index and absorption coefficient) and their derivatives, related to band structure and transport parameters (CP energies and broadenings), of thick Ge layers on Si are virtually identical to those of bulk Ge. These results are in agreement with RHEED and EPD counts. Therefore, we should expect that electronic and optoelectronic devices fabricated using Ge on Si should have similar (if not superior) characteristics compared to bulk Ge-based devices.

Ames Laboratory is operated for the U.S.-DOE by ISU under Contract No. W-7405-ENG-82. The work at Ames was supported by the Director of Energy Research, Office of BES, by the ISGC, and by NSF (DMR-9413492). The work at Delaware was supported by AFOSR (F49620-95-0135), ARO (DAAH04-95-1-0625), DARPA, and ONR (N00014-93-1-0393).

[1] B.-Y. Tsaur, M. W. Geis, J. C. C. Fan, and R. P. Gale, Appl. Phys. Lett. **38**, 779 (1981); J. M. Baribeau, T. E. Jackman, D. C. Houghton, P. Maingé, and M. W. Denhoff, J. Appl. Phys. **63**, 5738 (1988).

[2] D. P. Malta, J. B. Posthill, R. J. Markunas, and T. P. Humphreys, Appl. Phys. Lett. **60**, 844 (1992).

[3] S. S. Iyer, G. L. Patton, J. M. C. Stork, B. S. Meyerson, and D. L. Harame, IEEE Transactions on Electron Devices **36**, 2043 (1989).

[4] K. Ismail, F. K. LeGoues, K. L. Saenger, M. Arafa, J. O. Chu, P. M. Mooney, and B. S. Meyerson, Phys. Rev. Lett. **73**, 3447 (1994).

[5] B. A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J. W. Mayer, and J. Spear, Appl. Phys. Lett. **69**, 2557 (1996).

[6] K. L. Wang and R. P. G. Karunasiri, J. Vac. Sci. Technol. B **11**, 1159 (1993).

[7] P. Sheldon, B. G. Yacobi, K. M. Jones, and D. J. Dunlavy, J. Appl. Phys. **58**, 4186 (1985); R. D. Dupuis, J. C. Bean, J. M. Brown, A. T. Macrander, R. C. Miller, and L. C. Miller, J. Electron. Mater. **16**, 69 (1987).

[8] D. J. Eaglesham and M. Cerullo, Phys. Rev. Let **64**, 1943 (1990); D. J. Eaglesham and R. Hull, Mater. Sci. Eng. B **30**, 197 (1995).

[9] R. Hull and J. C. Bean, J. Vac. Sci. Technol. A **7**, 2580 (1989).

[10] E. A. Fitzgerald, Y.-H. Xie, M. L. Green, D. Brasen, A. R. Kortan, J. Michel, Y.-J. Mii, and B. E. Weir, Appl. Phys. Lett. **59**, 811 (1991); P. M. Mooney, J. L. Jordan-Sweet, J. O. Chu, and F. K. LeGoues, *ibid.* **66**, 3642 (1995).

[11] R. Lange, K. E. Junge, S. Zollner, S. S. Iyer, A. P. Powell, and K. Eberl, J. Appl. Phys. **80** 4578 (1996).

[12] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, Appl. Phys. Lett. **67**, 1865 (1995).

[13] D. E. Aspnes and A. A. Studna, Phys. Rev. B **27**, 985 (1983).

[14] J. L. Freeouf, J. C. Tsang, F. K. LeGoues, and S.S. Iyer, Phys. Rev. Lett. **64**, 315 (1990).

[15] J. E. Hulse and S. J. Rolfe, Thin Solid Films **93**, 223 (1993).

[16] P. Etchegoin, J. Kircher, M. Cardona, and C. Grein, Phys. Rev. B **45**, 11721 (1992).

[17] L. Viña, S. Logothetidis, and M. Cardona, Phys. Rev. B **30**, 1979 (1984).

Band-edge photoluminescence from pseudomorphic $Si_{0.96}Sn_{0.04}$ alloy

Al-Sameen T. Khan[a)] and Paul R. Berger[b)]
Department of Electrical Engineering, University of Delaware, Newark, Delaware 19716

Fernando J. Guarin[c)]
IBM Microelectronics Division, 1580 Route 52, Hopewell Junction, New York 12533

Subramanian S. Iyer[c)]
SiBond L.L.C., Hudson Valley Research Park, 1580 Route 52, Hopewell Junction, New York 12533

(Received 28 December 1995; accepted for publication 20 March 1996)

Band-edge related photoluminescence from a strained $Si_{0.96}Sn_{0.04}$ alloy grown by molecular beam epitaxy on Si(100) substrate has been seen for the first time. We report band-edge related photoluminescence from a compressively strained pseudomorphic $Si_{0.96}Sn_{0.04}$ alloy. The luminescence observed consisted of two dominant features, a well-resolved band-edge luminescence consisting of a no-phonon and a transverse optical phonon replica, and a deep-level broad luminescence peak around 770 meV. The band-edge feature is attributed to a no-phonon free excitonic recombination in the binary alloy and exhibits a near linear power dependence. We also observe a red shift of the energy gap of $Si_{0.96}Sn_{0.04}$ alloy with respect to Si, which corresponds to the bulk alloy effect. © *1996 American Institute of Physics.* [S0003-6951(96)01422-2]

Recently, there has been a lot of attention focused on the optical and electronic properties of group IV alloys grown on silicon.[1] The most studied system is the $Si_{1-x}Ge_x$ binary alloy[2–4] but recently the binary $Si_{1-x}C_x$,[5,6] $Ge_{1-x}C_x$,[7,8] the ternary $Si_{1-x-y}Ge_xC_y$ (Ref. 9) alloys, and quaternary alloys of $Si_{1-x-y-z}Ge_xC_ySn_z$ (Ref. 10) have received some attention as well. The binary IV–IV alloy $Si_{1-x}Sn_x$ has also been studied.[10,11] The majority of the previous investigations was concerned with the preparation and characterization of amorphous films, but recently there have been a few studies in the growth of crystalline thin films of $Si_{1-x}Sn_x$ and $Si_{1-x-y}Sn_xC_y$ alloys.[10,11] The amorphous films were prepared by magnetron sputtering,[12] rf sputtering,[13–15] chemical vapor deposition,[16] dc triode sputtering,[17] e-beam Knudsen cell evaporation,[18] and recently pulsed laser deposition,[19] with various degrees of success. The crystalline thin films have been prepared previously by a combination of ion implantation and solid-phase epitaxy (SPE).[20,21] This method of synthesis, however, was plagued with residual ion damage and the inability to produce sharp interfaces. Recently, high-quality pseudomorphic crystalline thin films of $Si_{1-x}Sn_x$ alloys have been prepared by molecular beam epitaxy (MBE).[10,11] These previous studies have all indicated that it is very difficult to stabilize Sn in the Si alloy lattice. This is mainly due to the following reasons. There is a large difference in the Si and Sn lattice parameters. Grey tin has a lattice parameter of 6.489 Å and Si a value of 5.431 Å, resulting in a 19.48% lattice mismatch. Grey tin (α-Sn) crystallizes in the diamond structure and at 13.2 °C it transforms into the tetragonal structure of metallic white tin (β-Sn). It is possible to incorporate a large concentration of Sn in a disordered Si:H matrix,[12,16] but Sn has a very low solid solubility in crystalline silicon[22] ($\sim 5\times 10^{19}$ cm^{-3}). The incorporation of isoelectronic Sn into silicon would result in the ability to engineer the band gap, which would lead to changes in the optical and electrical properties of the material. According to recent theoretical calculations,[23] $Si_{1-x}Sn_x$ alloy would ideally have a band gap between that of silicon and tin, 0.08 and 1.153 eV, which is very important for IR detector applications. Also, the band gap was predicted[23] to be direct for the compositional range of $0.9<x<1$ and indirect for all others. Strained layers of $Si_{1-x}Sn_x$ alloy semiconductors would find numerous applications in electronic and optoelectronic heterostructures. To our knowledge, the optical luminescence properties of pseudomorphic $Si_{1-x}Sn_x$ alloys grown on Si substrates have not been reported in any previous studies.

For the first time, we report in this letter the band-edge photoluminescence of compressively strained $Si_{1-x}Sn_x$ alloy grown by molecular beam epitaxy (MBE) on Si(100) substrate. The luminescence at 6 K consisted of a no-phonon (NP), a transverse optic (TO) phonon, and a deep-level broad luminescence peak. The band-edge feature is attributed to a no-phonon free excitonic recombination in the binary alloy and is found to exhibit a near linear power dependence. The band-gap energy variation of the alloy is determined from the position of the NP line. This is found to be in good agreement with the effect of Sn in the bulk band gap but appears to be redshifted due to residual strain.

The details of the solid source MBE growth of the $Si_{1-x}Sn_x$ layer are described elsewhere.[10] After an RCA type clean, the (100) silicon wafers were cleaned *in situ* at over 950 °C to desorb the chemical oxide. A thin 20 nm Si buffer layer was grown initially followed by the growth of the $Si_{1-x}Sn_x$ layer. The Si flux was feedback controlled and the Sn flux was controlled through the temperature control of the Knudsen cell. The $Si_{1-x}Sn_x$ layer growth temperature was 500 °C and the layer thickness was 150 nm. The growth conditions were optimized to maximize the incorporation of Sn on Si lattice sites. The layer composition were measured by Rutherford backscattering spectrometry (RBS) and con-

---

[a)] Permanent address: Delaware State University, Department of Physics & Astronomy, Dover, DE, 19901; Electronic mail: khan@ee.udel.edu

[b)] Electronic mail: pberger@ee.udel.edu

[c)] Department of Electrical Engineering, Columbia University, New York, NY 10027.

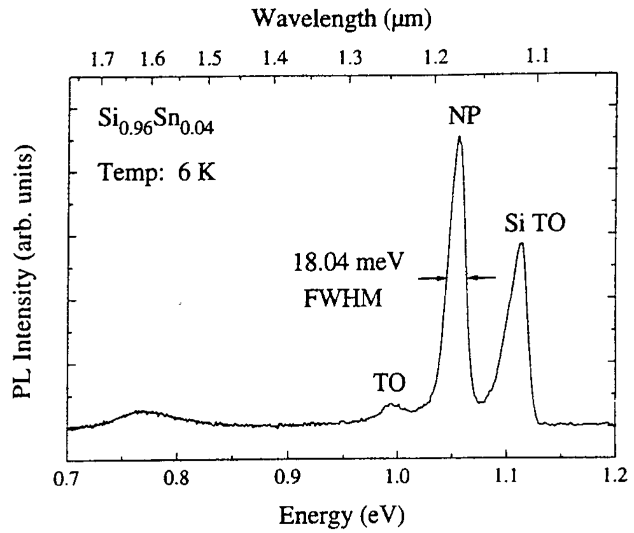


FIG. 1. PL spectra at 6 K of a $Si_{0.96}Sn_{0.04}$ alloy grown by MBE showing band-edge (NP) and deep-level luminescence. Laser excitation consisted of Ar^{+} ion laser with a laser spot size of 0.68 mm diameter.

firmed by x-ray diffraction analysis (XRD). X-ray diffraction measurements of the layer confirmed the substitutional incorporation of Sn and that the layer was pseudomorphic and compressively strained.[10] Alloy decomposition through precipitation of tin atoms into metallic β-Sn was not observed at a growth temperature of 500 °C, unlike that observed by Shiryaev *et al.*[11] after high-temperature (1000 °C) annealing.

Photoluminescence (PL) spectra were recorded in standard lock-in configuration, using a dispersive 1 m high-resolution Jarell–Ash (Czerny–Turner) monochromator and detected by a liquid-nitrogen-cooled Ge *p-i-n* photodetector (North Coast EO-817L). The samples were mounted on a cold finger in a temperature variable helium-flow cryostat. The excitation was provided by a multi-line cw Argon ion laser (488–514 nm) with pump intensities between 0.3 and 5 W/cm^2. Data collection and lock-in amplification were controlled by a desktop computer.

PL spectra for the $Si_{0.96}Sn_{0.04}$ sample is presented in Fig. 1. The luminescence consists of a low intensity deep-level luminescence and two band-edge peaks. The most intense peak at 1.055 eV is ascribed to a NP transition, whereas the second peak at 0.9938 eV is redshifted by 61 meV and corresponds to its transverse-optic (TO) Si–Si phonon replica. This NP transition is shifted 98 meV below that observed in pure Si.[24] We attribute this redshift to the reduction in band gap of the binary SiSn alloy. However, the observed NP transition is also shifted 54.8 meV below an expected 1.1106 eV luminescence position using Vegard's Law for a SiSn alloy composition with 4% Sn. This additional redshift we attribute to the reduction in the band gap caused by residual strain in the pseudomorphic layer. The observed luminescence peak at 1.114 eV is attributed to a Si transverse-optical (TO) phonon transition within the Si substrate. Also, a deep-level broadband luminescence is observed around 770 meV. Similar deep-level broadband luminescence has been observed in SiGe, SiC, and SiGeC layers grown by MBE and has been attributed to localized excitons in a strain field created by Ge platelets,[25] the emission of deep pseudoacceptors,[26] or to carbon-oxygen complexes,[5] respectively. It should also be noted that the samples were measured at various times over an eight-month period, and the luminescence spectra was very consistent.

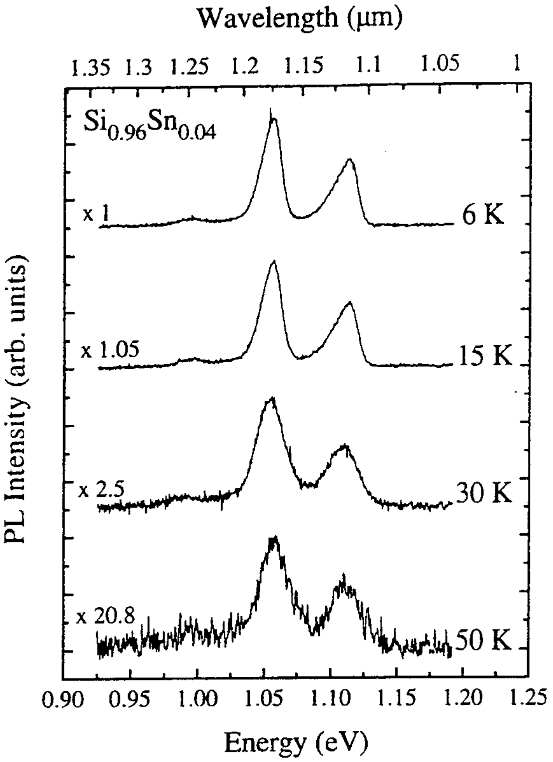


FIG. 2. PL spectra of $Si_{0.96}Sn_{0.04}$ alloy at 6, 15, 30, and 50 K.

The temperature dependence of the PL spectra was also measured. Temperature dependence of the PL spectra is shown in Fig. 2 by plotting the luminescence at 6, 15, 30, and 50 K. It was observed that the luminescence persisted past 50 K, but was degraded in intensity. As the temperature was increased the PL intensity decreased and the linewidth broadened in the direction of higher energy. This type of linewidth behavior is the characteristic of Maxwell–Boltzman distribution. The energy position of the NP peak remained constant as the temperature was increased. PL temperature dependence of this nature is the characteristic of free-exciton (FE) recombination. Therefore, we attribute the line labeled NP at 1.055 eV as due to a no-phonon free-exciton recombination. The full width at half maximum of the NP line is 18.04 meV at 6 K, which is rather broad compared to the intrinsic FE thermal linewidth, but this could be attributed to statistical fluctuations in the atomic distributions of the alloy as discussed by Robbins *et al.*[27]

The laser power dependence of the band-edge luminescence and the deep-level peaks are shown in Fig. 3. The peak labeled NP shows an almost linear increase in intensity with increasing laser power, while the deep-level broadband shows a sublinear dependence. Linear power dependence is a characteristic of a no-phonon free-exciton recombination. The square-root dependence of the deep-level band indicates that a recombination center is involved in the recombination process. This behavior is a characteristic of localized excitons bound to an isoelectronic trap in Si.[28]

In conclusion, we have reported band-edge related pho-

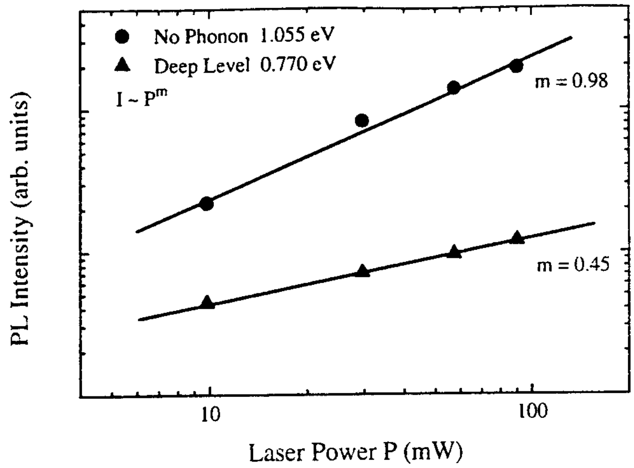


FIG. 3. Intensity if the NP transition and the deep-level transition for a $Si_{0.96}Sn_{0.04}$ alloy grown by MBE as a function of laser power.

toluminescence from a pseudomorphic $Si_{0.96}Sn_{0.04}$ alloy grown by MBE. Two general features have been observed for the first time, a deep-level broadband luminescence and a band-edge luminescence consisting of a NP and TO replica. The band-edge feature increases almost linearly with increasing laser power while the deep-level broadband showed a square-root dependence. The redshift of the energy gap was attributed to the bulk alloy effect and to residual strain.

[1] R. A. Soref, Proc. IEEE **81**, 1687 (1993).
[2] J. Weber and M. I. Alonso, Phys. Rev. B **40**, 5683 (1989).
[3] J. C. Sturm, H. Manoharan, L. C. Lenchyshyn, M. L. Thewalt, N. L. Rowell, J. P. Noël, and D. C. Houghton, Phys. Rev. Lett. **66**, 1362 (1991).
[4] J. Brunner, J. Nützel, M. Gail, U. Menczigar, and G. Abstreiter, J. Vac. Sci. Technol. B **11**, 1097 (1993).
[5] P. Boucaud, C. Francis, A. Larré, F. H. Julien, J.-M. Loutioz, D. Bouchier, S. Bodnar, and J. L. Regolini, Appl. Phys. Lett. **66**, 70 (1995).
[6] W. Kissinger, M. Weidner, H. J. Osten, and M. Eichler, Appl. Phys. Lett. **65**, 3356 (1994).
[7] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, Appl. Phys. Lett. **67**, 1865 (1995).
[8] H. J. Osten, E. Bugiel, and P. Zamuseil, J. Cryst. Growth **142**, 322 (1994).
[9] P. Boucaud, C. Francis, F. H. Julien, J.-M. Lourtioz, D. Bouchier, S. Bodnar, B. Lambert, and J. L. Regolini, Appl. Phys. Lett. **64**, 875 (1994).
[10] F. J. Guarin, S. S. Iyer, B. A. Ek, and A. R. Powell, Appl. Phys. Lett. **68** (1996) (to be published).
[11] S. Y. Shiryaev, J. L. Hansen, P. Kringhøj, and A. N. Larsen, Appl. Phys. Lett. **67**, 2287 (1995).
[12] Y. Kuwano, M. Ohnichi, and H. Nishiwaki, *Proceedings of the 16th IEEE Photovoltaics Specialist Conference, San Diego* (IEEE, New York, 1982), p. 1338.
[13] D. Girginoudi, N. Georgoulas, A. Thanailakis, A. Zoetsis, O. Kybriakidis, and A. Christou, Proc. Mater. Res. Soc. **118**, 703 (1988).
[14] D. L. Williamson, R. C. Kerns, and S. K. Deb, Appl. Phys. Lett. **55**, 2816 (1984).
[15] G. Chen, F. Zhang, and D. Ae, Sol. Energy Mater. **12**, 471 (1985).
[16] M. Vergnat, G. Marchal, M. Piecuch, and M. Gerl, Solid State Commun. **50**, 237 (1984).
[17] C. Verie, J. F. Rochette, and J. P. Rebouillat, J. Phys. (Paris) Colloq. **42**, Suppl. 10, C4-667 (1981).
[18] D. Girginoudi, A. Thanailakis, and A. Christou, Proc. Mater. Res. Soc. **77**, 603 (1987).
[19] R. E. Treece, J. S. Horwitz, D. B. Chrisey, J. Tang, and R. S. Williams, Proc. Mater. Res. Soc. **358**, 775 (1995).
[20] M. R. Sardela, R. Turan, M. Willander, G. V. Hansson, and L. Hultman, J. Appl. Phys. **77**, 1411 (1995).
[21] P. Mei, M. T. Schmidt, E. S. Yang, and B. J. Wilkens, J. Appl. Phys. **69**, 8417 (1991).
[22] F. A. Trumbore, C. R. Isenberg, and E. M. Porbansky, J. Phys. Chem. Solids **9**, 60 (1959).
[23] R. A. Soref and C. H. Perry, J. Appl. Phys. **69**, 539 (1991).
[24] P. J. Dean, J. R. Haynes, and W. F. Flood, Phys. Rev. **161**, 711 (1967).
[25] J. P. Noël, N. L. Rowell, D. C. Houghton, and D. D. Perovic, Appl. Phys. Lett. **61**, 690 (1992).
[26] J. Denzel, K. Thonke, J. Spitzer, R. Sauer, H. Kibbel, H.-J. Herzog, and E. Kasper, Thin Solid Films **222**, 89 (1992).
[27] D. J. Robbins, L. T. Canham, S. J. Barnett, A. D. Pitt, and P. Calcott, J. Appl. Phys. **71**, 1407 (1992).
[28] J. Weber, W. Schmid, and R. Sauer, Phys. Rev. B **21**, 2401 (1980).

ELSEVIER

Thin Solid Films 302 (1997) 201-203

thin solid films

Energy band offsets of SiGeC heterojunctions [1]

J. Kolodzey [a,*], F. Chen [a], B.A. Orner [a], D. Guerin [b], S. Ismat Shah [c]

[a] *Department of Electrical Engineering, University of Delaware, 140 Evans Hall, Newark, DE 19716, USA*
[b] *Department of Physics and Astronomy, University of Delaware, Newark, DE 19716, USA*
[c] *Dupont Experimental Station, Wilmington, DE 19898, USA*

Abstract

We report on conduction and valence band offsets in thick, relaxed Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys grown by solid source molecular beam epitaxy on (100) Si substrates. X-ray photoemission spectroscopy was used to measure the valence band energies with respect to atomic core levels, and showed that C increased the valence band maximum of SiGeC by +48 meV/%C. The bandgap energies were obtained from optical absorption, and were combined with the valence band offsets to yield the conduction band offsets. For SiGeC/Si heterojunctions, the offsets were typically 0.6 eV for the valence band and 0.38 eV for the conduction band, with a staggered type II alignment. These offsets can provide significant electron and hole confinement for device applications. © 1997 Elsevier Science S.A.

Keywords: Band structure; Heterostructures; Semiconductors; X-ray photoelectron spectroscopy (XPS)

---

1. Introduction

Alloys of $Si_{1-x-y}Ge_xC_y$ are exciting candidates for the heterojunction devices expected in future integrated circuits. The bandgaps and the energy offsets of the conduction and valence bands at heterointerfaces are important parameters for device design [1,2], but are not well known for the SiGeC system. Electrical measurements performed directly on heterojunctions may be misleading because of interface charges caused by defects and impurities. Although no measurement technique is without difficulties, it is generally acknowledged that X-ray photoelectron spectroscopy (XPS) gives unambiguous results [3–5].

We used XPS to measure the valence bands offsets that would exist in heterojunctions of $Si_{1-x-y}Ge_xC_y$ alloy layers. The energies of the valence band maxima (VBM) were measured with respect to the atomic core levels of Si and Ge. The conduction band offsets were determined by including the bandgap energies obtained from optical transmission measurements. The layers were grown by molecular beam epitaxy (MBE) on (100) Si substrates and were unusual in having Ge-rich compositions with up to 2 at% C. Single, relaxed epitaxial layers exceeding the critical thickness were measured to avoid ambiguities that may occur in pseudomorphic heterojunctions caused by interface defects and the degree of strain.

2. Experimental details

The MBE technique used to grow the $Si_{1-x-y}Ge_xC_y$ alloys was described elsewhere [6,7]. Solid thermal sources were used for the Si and Ge, and the C was sublimated from a resistively heated filament. Substrates were 75 mm diameter (100)-oriented Si wafers 750 μm thick. Substrates were degreased and etched ex-situ followed by an HF dip, and were immediately loaded into the MBE chamber for outgassing for 1 h at 250 °C. The SiGeC alloys were grown at a substrate temperature of 600 °C at a rate of 0.1 $\mu m\ h^{-1}$. Prior to growth, reflection high-energy electron diffraction (RHEED) showed the (2×1) reconstructed surface of Si, and after growth indicated flat crystalline layers. Visual inspection showed the layer surfaces were specular and smooth. X-ray diffraction indicated that the layers were single crystalline with (100) orientation.

The XPS system used an SSX-100 spectrometer with Al Kα radiation, as described elsewhere [4]. Pure Si, Ge and C (thin film diamond) samples were used for calibration. The relative intensities of the Si 2p, the Ge 3d, and the C 1s core levels indicated that the compositions agreed with Rutherford backscattering (RBS) measurements using the

---

[*] Corresponding author. Tel: +1 302 831 1164; fax: +1 302 831 4316; email: kolodzey@ee.udel.edu.

[1] This paper was presented as part of the EMRS 1996 Spring Meeting, Symposium D, Strasbourg, France, June 4–7, 1996.

0040-6090/97/$17.00 © 1997 Elsevier Science S.A. All rights reserved.
PII S0040-6090(97)00027-8

C resonance at 4.265 MeV to enhance the C signal [8].

The offsets were determined from the relative differences between the VBM and the Si 2p and the Ge 3d core levels. Atomic core level positions were determined by Gaussian fits to the intensities. Variations in chemical bonding with alloy composition may slightly shift the core levels for different samples. Analysis of $Ge_{1-y}C_y$ alloys indicated that the Ge 3d core level energy shifted by at most 900*y* meV for C fraction *y* [4]. Therefore, our assumption of fixed core levels may contribute 9 meV/%C of error. Additional sources of error include composition-dependent wavevector selection rules and interface effects [3].

The optical absorption coefficient and the bandgap energies were determined from optical transmission using Fourier transform infrared spectroscopy (FTIR), as described elsewhere [9]. Alloy measurements were referenced to blank Si substrates, and interference fringes were curve-fitted to determine the onset of fundamental absorption. Absorption coefficients were calculated by matching data to transmittance and reflectance equations.

3. Results and discussion

Fig. 1 shows the valence band energy distribution functions, proportional to the valence band density of states (DOS), for a SiGeC alloy and pure Si. The intensity peaks are related to critical points in the valence band DOS. Fig. 2 shows the valence band offset for a SiGeC alloy compared with Si. As a graphical construction, the offset between two compositions was determined from the amount of relative shift in the energy scales needed to align the valence band edges as in the Fig. 2 inset. This graphical

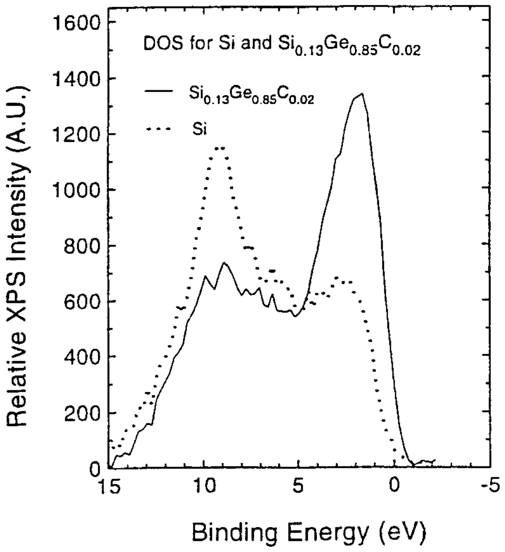


Fig. 1. Measured XPS intensity versus binding energy showing valence band energy distribution functions of $Si_{0.13}Ge_{0.85}C_{0.02}$ and of Si. Intensity peaks reflect variations in the valence band structure of the materials.

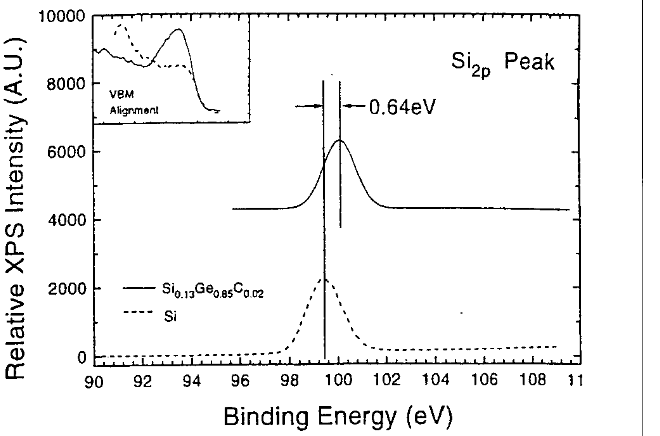


Fig. 2. Valence band offset energy between $Si_{0.13}Ge_{0.85}C_{0.02}$ and illustrating our technique of shifting the binding energy scales to align valence band edges as shown in the inset. The offset energy correspo to the amount of the shift, determined by comparing core level positio The energy scale is not absolute because it was shifted to align valence bands.

determination avoids errors that may occur when estim ing the VBM by extrapolating the band edge slopes to t different background intensities of each sample.

Table 1 summarizes the effects of C on the valen band offsets for SiGeC alloys with similar Si and contents but slightly different C fractions. To adjust for t different Si and Ge compositions between samples, used the published value of 7.8 meV/%Ge for the valen band offset [3], resulting in a VBM shift of +48 meV/% for our Ge-rich SiGeC alloys. The error per C percenta is estimated to include 4 meV due to uncertainties in t Ge fraction plus 9 meV due to chemical shifts. The offs reported here were not adjusted for strain because c alloys greatly exceed the critical thickness for dislocati formation, and the residual biaxial strain was measured X-ray diffraction to be less than 2×10^{-3}.

The measured optical absorption coefficient α is sho in Fig. 3 for a SiGeC alloy having a bandgap of 0.88 e For the SiGeC alloys measured here, α varied as t square of photon energy, indicating that the energy ga were indirect in *k*-space.

Table 1
Measured properties of SiGeC alloys including composition, band (E_g), valence band offsets (ΔE_v) and conduction band offsets (Δ with respect to Si and to Ge. Positive offset values indicate hig energies of the alloy band energies with respect to Si or Ge. Negat offsets indicate lower energies of the alloy bands

Sample	SGC-67	SGC-68	SGC-70
Composition	$Si_{0.13}Ge_{0.85}C_{0.02}$	$Si_{0.11}Ge_{0.88}C_{0.01}$	$Si_{0.10}Ge_{0.89}C_{0.}$
Thickness (μm)	0.22	0.18	0.19
E_g (eV)	0.889	0.866	0.854
ΔE_v on Si (eV)	+0.64	+0.62	+0.62
ΔE_c on Si (eV)	0.409	0.366	0.354
ΔE_v on Ge (eV)	+0.01	−0.01	0.00
ΔE_c on Ge (eV)	+0.229	+0.186	0.184

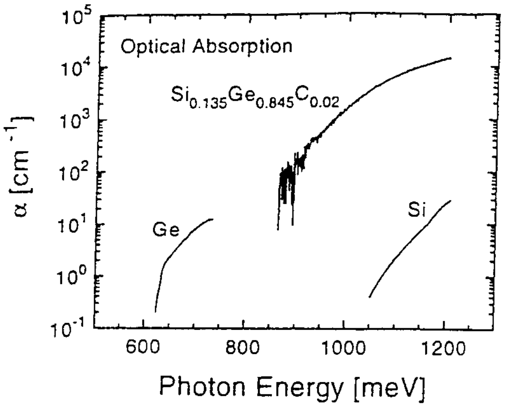


Fig. 3. Optical absorption coefficient versus photon energy of $Si_{0.13}Ge_{0.85}C_{0.02}$ and Si, showing the bandgap absorption edge. The bandgap of SiGeC lies below that of Si because of relatively high Ge content. Analysis indicated that alloy energy bands are indirect.

The conduction band offsets in Table 1 were obtained by adding the measured bandgaps to the valence band offsets. Our results show the band alignments to be staggered type II for SiGeC/Si heterostructures, as shown in the flat band diagram of Fig. 4. The valence band offsets for heterostructures of SiGeC alloys on Ge were much smaller than on Si. The valence band offset of pure Ge on Si can be obtained by subtracting the offsets for SiGeC on Ge from that of SiGeC on Si. The result for our samples yields 0.63 eV, which is smaller than the value of 0.78 eV reported for Ge on Si [3]. We ascribe this difference to experimental error, and to possible departures from transitivity [5].

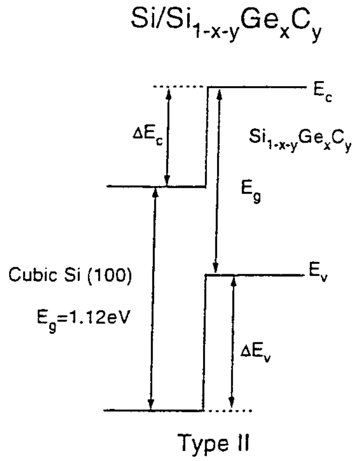


Fig. 4. Flat band diagram of SiGeC/Si heterojunction based on the data of Table 1. Energy band alignment is staggered, type II.

Using XPS, we have measured the valence band energies of thick bulk-like SiGeC alloys. Unlike previous studies, we report on Ge-rich compositions. We found that the addition of C shifts the VBM to higher energies. The conduction and valence band offsets between SiGeC and Si were large enough for significant confinement of both electrons and holes. This would not occur if the offsets were smaller than the thermal energy kT. Therefore SiGeC/Si heterojunctions can be useful for controlling both p-type and n-type conduction in device and circuit applications.

Acknowledgements

This work was supported by grants from the AFOSR (F49620-95-1-0135), ARO (DAAH04-95-1-0625), DARPA, and the ONR (N00014-93-1-0393). The authors gratefully acknowledge Drs T. Laursen and J. Mayer for the resonant RBS measurements, and J. Spear for X-ray diffraction measurements. Special thanks to Richard Soref for useful discussions and encouragement.

References

[1] A. St. Amour, C.W. Liu, J.C. Sturm, Y. Lacroix, M.L.W. Thewalt, *Appl. Phys. Lett.*, *67* (1995) 3915; 68 (1996) 1169.
[2] W. Kissinger, H.J. Osten, M. Weidner, M. Eichler, *J. Appl. Phys.*, *79* (1996) 3016.
[3] J.F. Morar, P.E. Batson, J. Tersoff, *Phys. Rev.*, *B 47* (1987) 4107.
[4] F. Chen, M.M. Waite, S. Ismat Shah, B.A. Orner, S.S. Iyer, J. Kolodzey, *Appl. Surf. Sci.*, *104/105* (1996) 615.
[5] G. Martin, A. Botchkarev, A. Rockett, H. Morkoc, *Appl. Phys. Lett.*, *68* (1996) 2541.
[6] J. Kolodzey, P.A. O'Neil, S. Zhang, B.A. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite, S. Ismat Shah, *Appl. Phys. Lett.*, *67* (1995) 1865.
[7] B.A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R.G. Wilson, P.R. Berger, J. Kolodzey, *J. Electronic Mater.*, *25* (1996) 297.
[8] B.A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J.W. Mayer, J. Spear, *Appl. Phys. Lett.*, *69* (1996) 2557.
[9] B.A. Orner, D. Hits, J. Kolodzey, F.J. Guarin, A.R. Powell, S.S. Iyer, *J. Appl. Phys.*, *79* (1996) 8656.

Appendix 24

IEEE ELECTRON DEVICE LETTERS, VOL. 18, NO. 1, JANUARY 1997 7

Low Resistance Ohmic Contacts to p-$Ge_{1-x}C_x$ on Si

Xiaoping Shao, S. L. Rommel, *Student Member, IEEE*, B. A. Orner, Paul R. Berger, *Member, IEEE*,
J. Kolodzey, *Senior Member, IEEE*, and K. M. Unruh

***Abstract*— We report on ohmic contact measurements of Al, Au, and W metallizations to p-type epitaxial $Ge_{0.9983}C_{0.0017}$ grown on a (100) Si substrate by molecular beam epitaxy (MBE). Contacts were annealed at various temperatures, and values of specific contact resistance have been achieved which range from $10^{-5}\ \Omega \cdot cm^2$ to as low as $5.6 \times 10^{-6}\ \Omega \cdot cm^2$. Theoretical calculations of the contact resistance of metals on $Ge_{1-x}C_x$ with small percentages of carbon, based on the thermionic field emission mechanism of conduction, result in good agreement with the experimental data. We conclude that Al and Au are suitable ohmic contacts to p-$Ge_{0.9983}C_{0.0017}$ alloys.**

I. INTRODUCTION

INVESTIGATION of Group IV alloys has been attracting more attention recently [1]. Previous Group IV alloy work has focused upon the use of lattice mismatched $Si_{1-x}Ge_x$ alloys on Si substrates to enhance device performance [2]. However, the $Si_{1-x-y}Ge_xC_y$ ternary alloy exhibits adjustable bandgaps while allowing the lattice constant to be varied from compressively to tensially strained about the Si lattice matching condition. For this reason, renewed attention has shifted to the $Si_{1-x-y}Ge_xC_y$ [3] and $Ge_{1-x}C_x$ [4] material systems. Mostly this research has centered upon material properties, but very recent reports have discussed $Ge_{1-x}C_x$ p-n [5] and $Si_{1-x-y}Ge_xC_y$ p-i-n [6] diodes. Related to device issues, the quality of the ohmic contacts need to be investigated.

Ohmic contacts are of great importance and an essential part of all solid-state device fabrication. No ohmic contact measurement to $Ge_{1-x}C_x$ alloys has been reported yet, however, and little work has been published for ohmic contacts on pure Ge either. The work reported here concentrates on Ge-rich $Ge_{1-x}C_x$ cubic heterostructures grown on Si by molecular beam epitaxy (MBE). In this paper, we present the first experimental results of metallization of the pure metals Al, Au, and W on epitaxial p-type $Ge_{1-x}C_x$ on Si for use as low resistance ohmic contacts.

II. EXPERIMENTS

The $Ge_{1-x}C_x$ epilayer was grown by solid source MBE in an EPI 620 system [7]. The substrate was (100) oriented n-type Si with a carrier concentration of 10^{15} cm^{-3}. The substrate temperature during growth was kept constant at 400 °C. The total thickness of the $Ge_{1-x}C_x$ epilayer was measured to be 0.6 μm. The $Ge_{1-x}C_x$ epilayer was doped p-type by a concurrent boron flux, using a third effusion cell loaded with pure boron in a pyrolytic graphite crucible. The Ge-rich layer was confirmed to be single crystal by X-ray analysis, but was relaxed as confirmed by transmission electron microscopy (TEM). From Hall effect measurements, the electrically active B concentration was about 4×10^{18} cm^{-3}, which was almost 100% activation, by comparing to the B concentration from secondary ion mass spectrometry (SIMS) measurements. The carbon concentration was 0.17%, as determined by the growth condition, calibrated from higher C concentrations.

Ohmic contacts to the p-type $Ge_{0.9983}C_{0.0017}$ were measured using the standard transmission line method (TLM) [8]. The TLM mesa patterns were fabricated by first defining rectangular GeC regions by photolithography and then etching the unprotected regions down to the Si substrate, using a H_3PO_4:H_2O_2:H_2O (1:6:3) etchant solution [9]. After degreasing, the contact area was defined by standard liftoff technology. Some samples received a brief etch prior to deposition, to remove about 300 Å of material and subsequent surface contamination. Al and Au contacts were evaporated thermally and by electron beam, respectively, while W contacts were magnetron sputtered. Each contact pad in the TLM pattern was 80 μm × 80 μm. The samples were then cleaved into smaller specimens, and underwent separate heat treatments at various temperatures in an annealing furnace under a forming gas (15% H_2–N_2) ambient. The annealing furnace consisted of a resitively heated graphite stage with a thermocouple embedded into the block. All annealing times were 3 min, the minimum response time of the annealing station, to achieve shallow ohmic contacts. Shallow contacts are needed for devices, such as bipolar transistors, to avoid diffusing through a p-n junction below.

III. RESULTS AND DISCUSSION

Fig. 1 shows a plot of the total resistance R_T measured between metal contacts as a function of the contact spacing between them. Three parameters, sheet resistance (ρ_s), contact resistance (R_c), and transfer length (L_T), were extracted from a least-squares interpolation line of the data. Assuming that the sheet resistance of the $Ge_{1-x}C_x$ epilayer outside the contact area (ρ_s) is the same as that beneath the contact area (ρ_{sc}), the total resistance is given by,

$$R_T = \frac{\rho_s d}{Z} + 2R_c \approx \frac{\rho_s d}{Z} + 2\frac{\rho_s L_T}{Z} \qquad (1)$$

Manuscript received August 6, 1996; revised October 2, 1996. This work was supported by AFOSR Grant F49620-95-1-0135 and DARPA (Sponsored Research Agreement with Texas Instruments, Inc.) under Grant SRA-3312665).

X. Shao, S. L. Rommel, B. A. Orner, P. R. Berger, and J. Kolodzey are with the Department of Electrical Engineering, University of Delaware, Newark, DE 19716 USA.

K. M. Unruh is with the Department of Physics, University of Delaware, Newark, DE 19716 USA.

Publisher Item Identifier S 0741-3106(97)00611-3.

0741–3106/97$10.00 © 1997 IEEE

Appendix 24

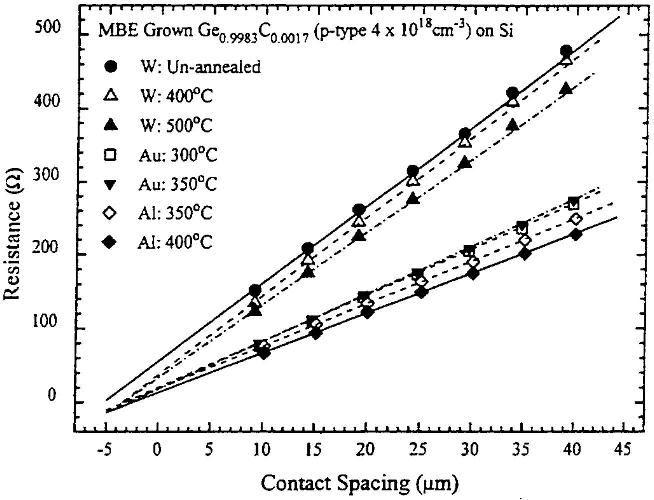


Fig. 1. The total resistance R_T measured between various metal contact pads in the TLM structure as a function of the contact separation for p-$Ge_{0.9983}C_{0.0017}$ on Si.

TABLE I
SPECIFIC CONTACT RESISTANCES EXTRACTED FROM THE DATA SHOWN IN FIG. 1 OF Al, Au, AND W METALS TO p-$Ge_{0.9983}C_{0.0017}$ ON Si UNDER VARIOUS ANNEAL TEMPERATURES

Contact Metal	Annealing Temperature (°C)	Specific Contact Resistance ρ_c (Ω·cm^2)
W	un-annealed	4.9×10^{-5}
W	400	2.0×10^{-5}
W	500	1.7×10^{-5}
Au	300	8.2×10^{-6}
Au	350	5.6×10^{-6}
Al	350	1.2×10^{-5}
Al	400	6.5×10^{-6}

where d is the contact spacing and Z is the contact width. The values of specific contact resistance, as shown in Table I, were calculated from these parameters.

Our results demonstrate low resistance ohmic contacts for Al and Au on $Ge_{0.9983}C_{0.0017}$, with values of specific contact resistance of the order of $10^{-5} - 10^{-6}$ $\Omega \cdot cm^2$, after suitable annealing. For the 300 °C anneal temperature, the Al contacts were found to be under-annealed, in which the plot of R_T versus d was an irregular scatter instead of a straight line. When the annealing temperature was raised to 450 °C or above, strong reactions of both Al and Au with the $Ge_{0.9983}C_{0.0017}$ epilayer occurred, which resulted in a highly irregular morphology with numerous metallic islands.

Optimal anneal temperatures based on resistances for Al and Au occurred within this temperature window, between these two extremes. The lowest contact resistances achieved were 6.5 x 10^{-6} $\Omega \cdot cm^2$ for Al at a 400 °C annealing temperature and 5.6 x 10^{-6} $\Omega \cdot cm^2$ for Au at 350 °C.

In spite of its higher contact resistance, tungsten was also explored as a nonalloyed contact and under heat treatment conditions. The W–GeC specific contact resistance was still on the order of 10^{-5} Ω·cm^2, even for an unannealed sample. The W contacts showed excellent adhesion to the $Ge_{0.9983}C_{0.0017}$, with a smooth surface even after successive temperature steps up to 650 °C. This indicated the low reactivity of W with GeC.

We consider three current transport mechanisms for a metal-semiconductor interface: thermionic emission (TE), thermionic field emission (TFE), and field emission (FE). The conduction of a metal-semiconductor contact is determined by the energy barrier height at the interface (Φ_B), the doping concentration near the semiconductor surface (N), the effective mass of the semiconductor majority charge carriers (m^*), the dielectric constant of the semiconductor (ϵ_s) and the temperature (T). Comparisons are given to the Mead rule using the Bardeen approximation [10], Tersoff's metal-induced gap states (MIGS) model [11], Tersoff's model based on a gap center and an adjustable parameter related to the metal electronegativity [12], and Cardona and Christensen's dielectric midpoint energy (DME) model [13].

The $Ge_{1-x}C_x$ alloy is expected to be covalently bonded and therefore have barrier heights independent of the metal work function for metals with weak chemical bonding. This is supported by the similar measured values of the specific contact resistance for Au and Al. Although it is known that barrier heights depend somewhat on the metal through its electronegativity [14], the agreement of contact resistances for Al and Au signifies a lack of dependence of barrier height upon type of metal, indicating that perhaps surface states and defects at the interface dominate.

To determine the conduction mechanism for the Au–GeC and Al–GeC (with small percentage of carbon) systems, we assumed Ge properties for the GeC alloy in the following calculation. This assumption is based on the small percentage of carbon used in this $Ge_{1-x}C_x$ study. For bulk Ge with a doping level of 4 x 10^{18} cm^{-3}, kT/E_{00} is 1.6, where

$$E_{00} = \frac{q\hbar}{2}\sqrt{\frac{N}{\epsilon_s m^*}} \tag{2}$$

is a characteristic energy. Based on this value of E_{00}, we expect the TFE mechanism to dominate, and the specific contact resistance can be calculated by [15], [16]

$$\rho_c = \frac{k}{qA^*T}\frac{kT}{\sqrt{\pi(\Phi_B + u_F)E_{00}}}\cosh\frac{E_{00}}{kT}\sqrt{\coth\frac{E_{00}}{kT}} \times \exp\left[\frac{\Phi_B + u_F}{E_0} - \frac{u_F}{kT}\right] \tag{3}$$

where A^* is the Richardson constant, u_F is the difference between the Fermi level and the valence band in semiconductor, and

$$E_0 = E_{00}\coth\frac{E_{00}}{kT}. \tag{4}$$

The values of specific contact resistance from the theoretical calculations, by using Mead, both Tersoff, and DME models

TABLE II

SPECIFIC CONTACT RESISTANCES FROM MEASUREMENTS AND THEORETICAL CALCULATIONS OF Al AND Au TO p-$Ge_{0.9983}C_{0.0017}$ ON Si

Metal	Specific Contact Resistance ρ_c ($\Omega\cdot cm^2$)					
	Our Data	Theoretical Values				
		Mead[10]	Tersoff (E_B)[11]	Tersoff (adjusted)[12]	DME[13]	Other[17]
Au	5.6×10^{-6}	6.4×10^{-6}	1.8×10^{-6}	9.7×10^{-8}	5.3×10^{-8}	7.6×10^{-8}
Al	6.5×10^{-6}	6.4×10^{-6}	1.8×10^{-6}		5.3×10^{-8}	1.8×10^{-6}

for the barrier height, are listed in Table II, and compared with our experimental data. As a comparison, values of specific contact resistance calculated from other workers' data [17] of barrier heights are also listed. The theoretical values impose a lower limit upon the ohmic contact measurements. Those theoretical values from Mead rule and Tersoff's effective midgap energy are very consistent with the experimental results obtained, while that from DME model gives a much lower limit.

Due to the high melting point of W, the heat of reaction of W-GeC is relatively large compared to Au or Al. This can be seen from our experiments by the fact that no morphological reaction was observed even after an anneal at 650 °C. It was expected that W could form a Schottky barrier on GeC due to the difference between the W work function and the electron affinity of $Ge_{1-x}C_x$ using small carbon percentages. However, reasonable ohmic contacts were achieved in this system. This is actually expected because the sputtering process which deposited the W metal probably produced defects sites near the semiconductor surface, thereby increasing the interface states density and lowering the barrier height.

IV. CONCLUSION

In conclusion, based on an industry standard of $10^{-5}\,\Omega\cdot cm^2$ [18], suitable low resistance ohmic contacts to p-$Ge_{0.9983}C_{0.0017}$ epitaxial layers grown on Si substrates using pure metallic contacts have been achieved. The contacts studied used Al, Au, and W metals. The metals Al and Au resulted in the lowest contact resistance of $6.5\times10^{-6}\,\Omega\cdot cm^2$ and $5.6\times10^{-6}\,\Omega\cdot cm^2$, respectively. The W metal also achieved reasonable results despite its low reactivity. Analysis implies a thermionic field emission model for the contact resistance.

ACKNOWLEDGMENT

The authors wish to thank M. McCarthy for W sputtering, T. Goodwin for Al deposition, R. G. Wilson for SIMS, and F. Chen for Hall measurements.

REFERENCES

[1] R. A. Soref, "Silicon-based optoelectronics," *Proc. IEEE*, vol. 81, pp. 1687–1706, 1993.

[2] G. L. Patton, J. H. Comfort, B. S. Meyerson, E. F. Crabbé, G. J. Scilla, E. De Frésart, J. M. C. Stork, J. Y.-C. Sun, D. L. Harame, and J. N. Burghartz, "75-GHz f_T SiGe-base heterojunction bipolar transistors," *IEEE Electron Device Lett.*, vol. 11, pp. 171–173, 1990.

[3] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, "Growth and strain compensation effects in the ternary $Si_{1-x-y}Ge_xC_y$ alloy system," *Appl. Phys. Lett.*, vol. 60, pp. 3033–3035, 1992.

[4] A.-S. T. Khan, X. Shao, B. A. Orner, P. R. Berger, and J. Kolodzey, "Photoluminescence of $Ge_{1-x}C_x$ alloys grown on Si (100) substrates," in *1996 MRS Spring Meet., Symp. F*, San Francisco, CA.

[5] X. Shao, S. L. Rommel, B. A. Orner, F. Chen, P. R. Berger, and J. Kolodzey, "Structural and electrical characterization of $Ge_{1-x}C_x$ cubic heterostructures on Si," in *1996 MRS Spring Meet., Symp. F*, San Francisco, CA.

[6] A. St. Amour and J. C. Sturm, "Pseudomorphic $Si_{1-x-y}Ge_xC_y$ pin diodes with low leakage current," in *1996 MRS Spring Meet., Symp. F*, San Francisco, CA.

[7] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, "Growth of germanium–carbon alloys on silicon substrates by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 67, pp. 1865–1867, 1995.

[8] D. K. Schroder, *Semiconductor Material and Device Characterization.* New York: Wiley, 1990, pp. 114–121.

[9] A.-S. T. Khan, "Photoluminescence and process issues of SiGeCSn alloys," Ph.D. dissertation, University of Delaware, 1996.

[10] C. A. Mead, "Metal-semiconductor surface barriers," *Solid-State Electron.*, vol. 9, pp. 1023–1033, 1966.

[11] J. Tersoff, "Schottky barrier heights and the continuum of gap states," *Phys. Rev. Lett.*, vol. 52, pp. 465–468, 1984.

[12] ——, "Schottky barriers and semiconductor band structures," *Phys. Rev. B*, vol. 32, pp. 6968–6971, 1985.

[13] M. Cardona and N. E. Christensen, "Acoustic deformation potentials and heterostructure band offsets in semiconductors," *Phys. Rev. B*, vol. 35, pp. 6182–6194, 1987.

[14] M. Schlüter, "Chemical trends in metal-semiconductor barrier heights," *Phys. Rev. B*, vol. 17, pp. 5044–5047, 1978.

[15] C. R. Crowell and V. L. Rideout, "Normalized thermionic-field (T-F) emission in metal-semiconductor (Schottky) barriers," *Solid-State Electron.*, vol. 12, pp. 89–105, 1969.

[16] A. Y. C. Yu, "Electron tunneling and contact resistance of metal-silicon contact barriers," *Solid-State Electron.*, vol. 13, pp. 239–247, 1970.

[17] S. M. Sze, *Physics of Semiconductor Devices.* New York: Wiley, 1981, p. 291.

[18] E. H. Rhoderick and R. H. Williams, *Metal-Semiconductor Contacts.* New York: Oxford, 1988, pp. 204–205.

ELSEVIER

Thin Solid Films 294 (1997) 122–124

thin solid films

Near band edge photoluminescence from pseudomorphic tensially strained $Si_{0.985}C_{0.015}$ alloy

Al-Sameen T. Khan [a,*], Paul R. Berger [a], Fernando J. Guarin [b], Subramanian S. Iyer [c]

[a] *Department of Electrical Engineering, University of Delaware, Newark, DE 19716, USA*
[b] *IBM Microelectronics Division, 1580 Route 52, Hopewell Junction, New York, NY 12533, USA*
[c] *SiBond L.L.C., Hudson Valley Research Park, 1580 Route 52, Hopewell Junction, New York, NY 12533, USA*

Abstract

Band edge related low temperature photoluminescence of a strained $Si_{0.985}C_{0.015}$ bulk alloy layer grown by molecular beam epitaxy on a Si (100) substrate has been investigated. The high quality layer was grown to a thickness of 1500 Å and was found to be pseudomorphic and tensially strained. We report two dominant features, a well-resolved band-edge luminescence consisting of a no-phonon and a transverse optical phonon replica, and an intense deep level luminescence peak around 0.778 eV. The band edge feature is attributed to a no-phonon free excitonic recombination in the binary alloy. We also observe a red shift of the energy gap of $Si_{0.985}C_{0.015}$ alloy with respect to Si, contrary to what is predicted according to the bulk alloy effect.

Keywords: Alloys; Molecular beam epitaxy; Photoluminescence

---

1. Introduction

Alloys of group IV elements have been under investigation recently as potential material systems for novel Si-based heterostructure device applications [1]. The $Si_{1-x}Ge_x$ material system [2,3] has received the most attention, with lesser efforts directed towards the growth and characterization of Sn-based [4] and C-based alloy layers. Lately, the C-based alloy systems of $Ge_{1-x}C_x$ [5,6], $Si_{1-x-y}Ge_xC_y$ [7], and $Si_{1-x-y-z}Ge_xC_ySn_z$ [8] have been receiving attention, especially $Si_{1-x}C_x$ [9,10]. However, in the past few years there has also been a number of investigations into the growth and characterization of $Si_{1-x}C_x$ alloy [11,12]. Two growth processes have mainly been used to grow $Si_{1-x}C_x$ alloys, chemical vapor deposition (CVD) [9,13], and molecular beam epitaxy (MBE) [14]. These two processes have been used to synthesize crystalline $Si_{1-x}C_x$ alloys and heterostructures containing a few percent of carbon. They indicate that only $Si_{1-x}C_x$ alloys which contain a few percent of C can be grown pseudomorphically on a Si substrate. This can be directly attributed to the following. Carbon has an extremely low solubility in Si, about 10^{-4} at.% at 1420°C, the melting point of silicon [15]. Additionally, there is a 40% mismatch in the bond lengths of Si and C. The mismatch would result in large local strain around the substitutional carbon atoms. Also, according to the Si–C phase diagram stoichiometric SiC is the only stable compound. Therefore, the substitutional incorporation of carbon above the solid solubility limit requires far from thermodynamic equilibrium growth conditions, such as MBE. The incorporation of isoelectronic carbon into silicon will offer new possibilities, for bandgap engineering, silicon compatible wide bandgap material, and large band offsets.

There have been a few theoretical investigations on the bandgap and electronic structure of $Si_{1-x}C_x$ alloys, but these studies come to conflicting conclusions. Soref [16], uses Vegards law and performs an interpolation giving a monotonic variation in the bandgap of $Si_{1-x}C_x$ alloys, between the constituent elements. Ab initio calculations were made by a some groups [17,18], revealing that for the addition of small carbon concentrations the $Si_{1-x}C_x$ bandgap reduces below Si, turns semi-metallic around 10% C, and then finally increases above Si beyond 10% C.

2. Experimental details

The details of the solid source MBE growth of the $Si_{1-x}C_x$ layer are described elsewhere [14]. The samples were grown using solid source MBE. Silicon was evaporated using a conventional electron beam evaporator and the carbon co-evaporated using a high purity pyrolytic graphite filament

---

[*] Corresponding author.

0040-6090/97/$17.00 © 1997 Published by Elsevier Science S.A.
PII S0040-6090(96)09460-6

housed in a water cooled source assembly. The C flux is controlled using the heating current through the filament. Calibration of the heating current was established using secondary ion mass spectroscopy (SIMS). The substrate temperature was maintained at 500–600°C depending on the carbon concentration of the layer grown. A thin 20 nm Si buffer layer was grown initially followed by the growth of the $Si_{1-x}C_x$ layer. By increasing the carbon concentration while growing an epitaxial silicon layer a tensially strained pseudomorphic $Si_{0.985}C_{0.015}$ alloy layer was obtained. The $Si_{0.985}C_{0.015}$ layer was grown to a thickness of 150 nm, and the growth conditions were optimized to maximize the incorporation of C on Si lattice sites. The layer composition and strain were measured by Rutherford backscattering spectrometry (RBS) and confirmed by X-ray diffraction analysis (XRD). X-Ray analysis confirmed the layer to be pseudomorphic, tensially strained, and the carbon was substitutionally incorporated in the lattice.

Photoluminescence spectra were recorded in standard lock-in configuration, using a dispersive 1 m high-resolution Jarell-Ash (Czerny–Turner) monochromator and detected by a liquid-nitrogen cooled Ge p–i–n photodetector (North Coast EO-817L). The samples were mounted on a cold finger in a temperature variable helium-flow cryostat. The excitation was provided by a multi-line cw argon laser (488–514 nm) focused to a sample area of approximately 2 mm^2, with pump intensities between 0.3 and 5 W cm^{-2}. Data collection, and lock-in amplification were controlled by a desktop computer.

3. Results and discussion

To the best of our knowledge, we know of only two other reports of PL from $Si_{1-x}C_x$ alloys. Boucaud et al. [9] reported PL from a $Si_{1-x}C_x$ bulk alloy layer grown by RTCVD which consisted of a deep level luminescence and was attributed to defects. Recently, Brunner et al. [19] reported near band edge luminescence from pseudomorphic $Si_{1-x}C_x$/Si multiple quantum well structures. Our PL measurements of bulk $Si_{1-x}C_x$ reveal both band edge luminescence and deep level luminescence.

Low temperature PL spectra at 5 K for the $Si_{0.985}C_{0.015}$ bulk alloy sample is presented in Fig. 1. The luminescence consists of an intense deep-level luminescence and two edge peaks. The most intense peak at 1.0599 eV is ascribed to a no-phonon (NP) transition whereas the second peak at 1.0019 eV is red-shifted by about 58 meV and corresponds to its transverse optic (TO) Si–Si phonon replica. The NP transition is shifted 93 meV below that observed in pure Si [20], and is additionally shifted about 159 meV below an expected 1.2182 eV luminescence position using Vegard's Law for a $Si_{1-x}C_x$ alloy composition with 1.5% C. We attribute this red shift to both the lattice distortion within the alloy due to highly localized strain around the carbon atoms, and residual strain between the pseudomorphic alloy layer and the substrate. The lowering of the bandgap due to the pseudomorphic strain energy in the layer is not enough to explain the large red shift alone. The observed luminescence peak at 1.1089 eV is attributed to a Si TO phonon transition within the Si substrate. Also, an intense deep-level broad band luminescence is observed around 0.7776 eV. Similar deep-level broad band luminescence has been observed in $Si_{1-x}Ge_x$ [21], $Si_{1-x-y}Ge_xC_y$ [7], $Si_{1-x}C_x$ [9], and $Ge_{1-x}C_x$ [23] layers grown by MBE, and were attributed to localized excitons in a strain field created by Ge platelets [21], the emission of deep pseudo-acceptors [22], or to carbon–oxygen complexes [9,23], respectively.

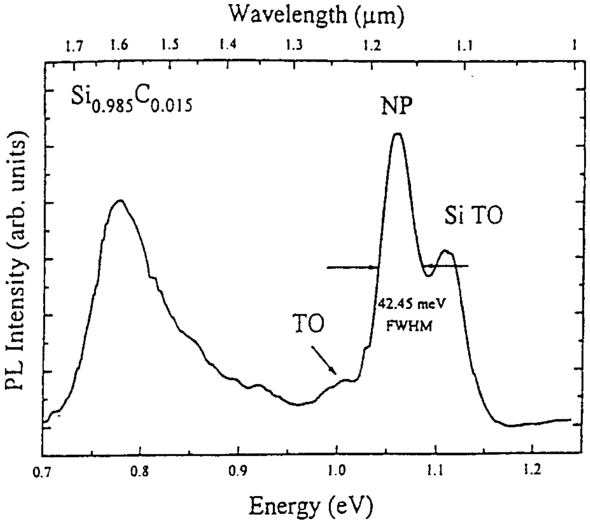


Fig. 1. PL spectra at 5 K of a 1500 Å $Si_{0.985}C_{0.015}$ alloy layer grown by MBE on (100) silicon substrate showing band-edge (NP) and deep level luminescence.

Temperature dependence of the PL spectra was also measured. The luminescence of the deep level peak persisted up to 250 K but was degraded in intensity, while the luminescence of the high-energy PL peaks persisted up to room temperature. As the temperature was increased, the PL intensity decreased, the peak position shifted linearly to higher energy, and the linewidth broadened in the direction of higher energy. This type of linewidth behavior is characteristic of the Maxwell–Boltzman distribution. The laser power dependence of the band-edge luminescence at 5 K show a linear increase in PL intensity with increasing excitation. Linear power dependence is a characteristic of a no-phonon free-exciton recombination. The deep-level band exhibited an almost linear increase in intensity with increasing laser power at low excitation levels, but as the power was increased to higher excitation levels, the luminescence saturates and a sub-linear almost square-root dependence is observed. This is indicative that a recombination center may be involved in the recombination process, and similar deep-level luminescence have previously been attributed to carbon–oxygen complexes [9,23].

4. Conclusions

In conclusion, we have reported low temperature photoluminescence of a pseudomorphic tensially strained

$Si_{0.985}C_{0.015}$ alloy layer grown by solid source MBE. Two general features have been observed, a deep-level broad band luminescence and a band-edge luminescence consisting of a NP and TO replica. Temperature dependent PL analysis of the band-edge feature indicated an exponential temperature dependence, characteristic of free excitonic recombination. The band-edge feature exhibited a linear power dependence with increasing laser excitation. The high-energy PL peaks persisted up to room temperature while the deep level peak persisted up to about 250 K but was severely degraded in intensity. The red-shift of the energy gap was attributed to lattice distortion within the alloy and residual strain between the alloy layer and the substrate.

References

[1] R.A. Soref, *Proc. IEEE, 81* (1993) 1687.

[2] J. Weber and M.I. Alonso, *Phys. Rev. B, 40* (1989) 5683.

[3] J.C. Sturm, H. Manoharan, L.C. Lenchyshyn, M.L. Thewalt, N.L. Rowell, J.P. Noël and D.C. Houghton, *Phys. Rev. Lett., 66* (1991) 1362.

[4] A.T. Khan, P.R. Berger, F.J. Guarin and S.S. Iyer, *Appl. Phys. Lett., 68* (1996) 3105.

[5] J. Kolodzey, P.A. O'Neil, S. Zhang, B.A. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite and S. Ismat Shah, *Appl. Phys. Lett., 67* (1995) 1865.

[6] H.J. Osten, E. Bugiel and P. Zamuseil, *J. Cryst. Growth, 142* (1994) 322.

[7] P. Boucaud, C. Francis, F.H. Julien, J.-M. Lourtioz, D. Bouchier, S. Bodnar, B. Lambert and J.L. Regolini, *Appl. Phys. Lett., 64* (1994) 875.

[8] F.J. Guarin, S.S. Iyer, B.A. Ek and A.R. Powell, *Appl. Phys. Lett.*, in press.

[9] P. Boucaud, C. Francis, A. Larré, F.H. Julien, J.-M. Lourtioz, D. Bouchier, S. Bodnar and J.L. Regolini, *Appl. Phys. Lett., 66* (1995) 70.

[10] W. Kissinger, M. Weidner, H.J. Osten and M. Eichler, *Appl. Phys. Lett., 65* (1994) 3356.

[11] H. Rücker, M. Methfessel, E. Bugiel and H.J. Osten, *Phys. Rev. Lett., 72* (1994) 3578.

[12] A.R. Powell, K. Eberl, F.E. LeGoues, B.A. Ek and S.S. Iyer, *J. Vac. Sci. Technol., B11* (3) (1993) 1064.

[13] J. Posthill, R. Rudder, S. Hattangady, G. Fountain and R. Markunas, *Appl. Phys. Lett., 56* (1990) 734.

[14] S.S. Iyer, K. Eberl, M.S. Goorsky, F.K. LeGoues, J.C. Tsang and F. Cardone, *Appl. Phys. Lett., 60* (1992) 356.

[15] R. Olesinski and G. Abbaschian, *Bull. Alloy Phase Diag., 5* (1984) 486.

[16] R.A. Soref, *J. Appl. Phys, 70* (1991) 2470.

[17] A.A. Demkov, and O.F. Sankey, *Phys. Rev. B, 48* (1993) 2207.

[18] J. Xie, K. Zhang and X. Xie, *J. Appl. Phys., 77* (1995) 3868.

[19] K. Brunner, K. Eberl and W. Winter, *Phys. Rev. Lett., 76* (1996) 303.

[20] P.J. Dean, J.R. Haynes and W.F. Flood, *Phys. Rev., 161* (1967) 711.

[21] J.-P. Noël, N.L. Rowell, D.C. Houghton and D.D. Perovic, *Appl. Phys. Lett., 57* (1990) 1037.

[22] J.-P. Noël, N.L. Rowell, D.C. Houghton and D.D. Perovic, *Appl. Phys. Lett., 61* (1992) 690.

[23] Khan et al., *Appl. Phys. Lett.*, submitted.

Appendix 26

$Si_{1-x-y}Ge_xC_y$ alloy band structures by linear combination of atomic orbitals

B. A. Orner[a] and J. Kolodzey
Department of Electrical Engineering, University of Delaware, Newark, Delaware 19716

(Received 24 October 1996; accepted for publication 7 February 1997)

We have applied a virtual crystal approximation to the linear combination of atomic orbitals method to calculate critical point energies of unstrained $Si_{1-x-y}Ge_xC_y$ alloys spanning the composition parameter space. Additionally, we have calculated the band structure across the Brillouin zone for a series of alloy compositions. We found the band energies had significant bowing departures from linearity throughout the system. In some cases, the energy band gap was not monotonically dependent on composition. Our theoretical results are compared with recent experimental results, and good agreement was found overall. © *1997 American Institute of Physics.* [S0021-8979(97)05010-X]

I. INTRODUCTION

The $Si_{1-x-y}Ge_xC_y$ alloy system is being investigated for use in group IV heterostructure devices.[1–5] At certain compositions, the system is lattice matched to Si, and could potentially be lattice matched to 3C-SiC. It could further be useful in lattice mismatched devices fabricated on Ge or diamond substrates, for example, or as a virtual substrate for other materials.

Usually, the ultimate motivation to explore new materials is to exploit a property of an alloy band structure that is not attainable using pure elements, such as band offsets in a lattice matched heterostructure, or an intrinsic property of the band structure such as lighter carrier effective mass. The band structure of group IV alloys, however, is largely unknown.

Because alloys containing C are metastable,[6] they cannot be produced by simple high temperature mixing. Several groups have recently produced crystalline random alloys with under 5% C using techniques such as molecular beam epitaxy or chemical vapor deposition.[7–9] Even for these C lean binary and ternary alloys, experimental band gap data are sparse. As the alloy materials are difficult to produce and analyze, a comprehensive experimental study across the whole composition range has not yet been completed.

For this reason, theoretical predictions of alloy band structure are useful first steps toward device design. Theoretical studies of the band structure of this alloy system have focused on $Si_{1-x}Ge_x$ and $Si_{1-y}C_y$ binary alloys, though a few ternary alloys have been studied.[10–12] Linear interpolation of critical points using the elemental values as endpoints yields a straightforward description of the band gap of any particular composition.[12] First principle calculations of $Si_{1-y}C_y$ alloys[10,13,14] have been incongruent with such a linear interpolation, predicting decreases in the band gap as C content is increased, and semimetallic materials at some compositions. These methods predicted the magnitude of this decrease to be between -20 and -100 meV/% C.

The first principles nature of these calculations is appealing, and the intriguing results have increased interest in the band structure of these alloys. These previous calculations, however, performed by the plane wave techniques,[13] *GW*,[14] and linear muffin tin orbital (LMTO)[10] methods, do have some drawbacks.

For example, previous calculations employed a supercell approach in which individual Si, Ge, and C atoms in appropriate proportions were located on diamond lattice sites to simulate the random alloy. The alloy composition is thus limited to discrete points in composition parameter space. The composition interval spacing is limited by the number of atoms in the supercell (in the referenced work, interval spacing was usually 12.5%). Our approach employs a virtual crystal approximation (VCA) within the linear combination of atomic orbitals (LCAO) method. This permits band structure calculations for alloys of arbitrary composition. A similar approach has been applied to $Si_{1-x}Ge_x$ alloys[15,16] but to our knowledge has not been applied to group IV semiconductor alloys containing C. As a future research subject, empirical pseudo-potential band structure calculations could be performed by a similar interpolation of parameters.

Furthermore, due to the fairly formidable calculations, only a few compositions were reported, primarily Si rich $Si_{1-y}C_y$ alloys. The LCAO calculations employed herein were accomplished for 1250 compositions.

Finally, the calculations generally (except for the *GW*[14] results) fail to accurately predict experimental band gap energies at the elemental end points, and are therefore questionable for conclusions on alloy properties. They are most useful in predicting trends in the band gap as composition is varied. To obtain absolute values for the band gap energies, an *ad hoc* correction factor must be included. If this correction factor is linearly interpolated to obtain alloy corrections, departures from linearity could give unexpected results. Due to the semiempirical nature of the LCAO method, the band structure at the elemental endpoints agree well with experimental values, and no correction factors were required.

II. THEORETICAL APPROACH

Our approach, like all methods, involves some assumptions and limitations. The VCA technique we employ does not account for localized strain effects resulting from a disparity in bond lengths between Si, Ge, and C. The smaller C atomic radius is expected to alter both the bond angles and lengths, resulting in a deviation from the ideal diamond crystal structure. Previous articles have reported that this departure decreases the energy band gap of these alloys.[11,13] The extent to which this effect decreases the band gap is not established. In this article we demonstrate that a VCA produces a large bowing departure in the band gap-composition relation. Combined with other studies, this may clarify the extent to which localized strain decreases the band gap.

---

[a] Electronic mail:orner@ee.udel.edu

Similarly, the VCA does not account for band broadening due to alloy disorder. This broadening can reduce the effective band gap energy and other transition energies. Techniques such as the coherent potential approximation (CPA)[17–19] do account for alloy disorder. The role of alloy disorder in $Si_{1-x-y}Ge_xC_y$ alloys has not been established. In the $Si_{1-x}Ge_x$ system, CPA studies[18,19] have found that disorder reduced the band gap energy by at most 7 meV, thus the CPA band gap energy differed only slightly from the VCA band gap energy. These studies did find larger band broadening further from the band edge, however, and cautioned that the extent of broadening could differ for other alloy systems. The VCA results presented here provide an initial point of information on the band structure of a wide range of $Si_{1-x-y}Ge_xC_y$ alloys. Comparison of these VCA results with future CPA studies will be of value in studies of alloy disorder band broadening in $Si_{1-x-y}Ge_xC_y$.

Therefore, we feel there is merit to exploring the band structure of this alloy system by several methods, and the approach taken herein is intended to complement previous work. This will allow for comparison between the results of a wide variety of different techniques. We do not claim the approach we employ is superior, but rather that, in light of sparse and inconsistent experimental data, many alternative theoretical and experimental approaches should be explored.

The LCAO formalism employed here is a modified version[20] of the classic work of Slater and Koster.[21] We use a three-center, third nearest-neighbor Hamiltonian. Traditionally, an orthogonal hybridized atomic orbital basis is assumed, with any non-orthogonality being absorbed by the empirical Slater-Koster (SK) parameters. We have performed our calculations by this traditional orthogonal LCAO method, but have also repeated some calculations using a non-orthogonal LCAO method.[22] For this, we formulated a separate overlap matrix with a separate set of empirical parameters to account for non-orthogonality. The Schroedinger equation in the non-orthogonal case is solved by

$$|S^{-1/2}HS^{-1/2}-EI|=0, \tag{1}$$

where H is an 8×8 Hamiltonian parameter matrix, S is an 8×8 overlap matrix, and E is the set of eigenvalues which solve the Schroedinger equation. In the orthogonal LCAO formulation, S is taken as the identity matrix. We focus initially and primarily on the results from the orthogonal method, due to the smaller parameter set involved. Limited results from the non-orthogonal method are also presented. The Hamiltonian matrix elements for the diamond structure, including third nearest neighbors[20,21] are listed in Table I.

Twenty empirical parameters were required for the orthogonal formulation, and 38 parameters were required for the non-orthogonal formulation. For the elemental end points we used those of Papaconstantopoulos,[20] which were obtained by curve fitting to pseudo-potential band structures.

Careful consideration must be given to the size of the empirical parameter set. A large parameter set permits excellent agreement between the band structure calculations and experimental results for the elemental end points. However, with larger parameter sets, it becomes difficult to obtain physically meaningful values for the SK parameters through curve fitting. This limits the confidence in the interpolated values to some extent—presumably any decrease in accuracy would be most pronounced for alloys in which no single element is the dominant component. We feel this fact is mitigated by the improved accuracy at and near the end points, and have thus employed larger parameter sets. Normally, accuracy is rapidly lost if fewer than 20 parameters are employed. By using a 10×10 matrix, Vogl *et al.*[23] have reduced the number of parameters to eight and have achieved good results, though the formulation we have employed still gives an improved representation of the band structure as evidenced by the superior experimental agreement of energies at the critical points.

The alloy SK parameters were obtained by interpolation. using the elemental SK parameters as end points. The on-site SK parameters were linearly interpolated with composition. The parameters for the first, second, and third nearest neighbors (for S as well as H) were interpolated in accordance with Harrison's d^{-2} rule:[23,24]

$$E_{\text{alloy}}=\left(\sum_{i=\text{Si,Ge,C}} d_i x_i\right)^{-2} \sum_{i=\text{Si,Ge,C}} (d_i)^2 x_i E_i, \tag{2}$$

where E_{alloy} and E_i are a particular empirical SK parameter for the alloy and elemental end points, respectively, and d_i is the appropriate nth nearest neighbor distance in crystalline Si, Ge, or C. Finally, x_i is the atomic fraction of each alloy component. The first term represents the nth nearest neighbor distance in the alloy. Here we have assumed Vegard's law is followed for these alloys, and have obtained the neighbor distances by linear interpolation.

After obtaining the set of SK parameters for an alloy, the eigenvalues were found by diagonalizing and solving Eq. (1) for a given wave vector **k**. To study the composition dependence of the critical points, the eigenvalues were obtained at the Γ and L symmetry points, and for 15 values of **k** on the Δ axis between $(2\pi/a)(0.7)[001]$ and $(2\pi/a)(0.95)[001]$ to obtain the Δ minimum. For selected alloys, the band structure was obtained by evaluating Eq. (1) on a grid of 100 **k** values.

III. RESULTS AND DISCUSSION

Using the orthogonal method, the composition dependence of the Γ, Δ, and L conduction band minima for each of the unstrained binary systems is indicated in Figs. 1–3. Nonlinear behavior was observed for all three binary systems, but was not pronounced for the $Si_{1-x}Ge_x$ system, for which we have compared the theoretical band gap to the

Appendix 26

TABLE I. Matrix elements for the H and S matrices. The H and S matrix elements have the same form and differ only by their empirical constants. On-site, first, second, and third nearest neighbors of the diamond lattice structure are included. Matrix elements H are described by a set of orbitals (s, x, or y) and numbers indicating the atomic site on the two-atom diamond basis. Wave vector $\mathbf{k}$ is expressed in units of $2\pi/a$, where a is the alloy lattice constant. Each empirical parameter E is designated by two letters indicating the two hybridized atomic orbitals of $\langle\Psi_1|H|\Psi_2\rangle$ and a number indicating the distance and direction (where distinct) between the atomic centers. Here sx indicates an s orbital and any p orbital (x, y, or z) xx indicates two p orbitals of equal magnetic quantum numbers, and xy indicates two p orbitals with differing magnetic quantum numbers.

$$\begin{aligned} H_{s1/s1}=H_{s2/s2}={}&E_{ss}(000)+4E_{ss}(220)[\cos(k_x)\cos(k_y)+\cos(k_y)\cos(kz)\\ &+\cos(k_x)\cos(k_z)]. \end{aligned}$$

$$\begin{aligned} H_{x1/x1}=H_{x2/x2}={}&E_{xx}(000)+4E_{xx}(220)[\cos(k_x)\cos(k_y)+\cos(k_x)\cos(k_z)]\\ &+4E_{xx}(022)\cos(k_y)\cos(k_z). \end{aligned}$$

$$\begin{aligned} H_{s1/s2}=H^*_{s2/s1}={}&4E_{ss}(111)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)\\ &+4E_{ss}(311)\left[\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)+\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_z}{2}\right)+\cos\left(\frac{k_z}{2}\right)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\right]\\ &-i4E_{ss}(111)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)\\ &+i4E_{ss}(311)\left[\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)+\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_z}{2}\right)+\sin\left(\frac{k_z}{2}\right)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\right]. \end{aligned}$$

$$\begin{aligned} H_{s1/x2}=-H^*_{s2/x1}={}&4E_{sx}(111)\cos\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)\\ &-4E_{sx}(311)\cos\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)\\ &-4E_{sx}(113)\cos\left(\frac{k_x}{2}\right)\left[\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)+\sin\left(\frac{k_z}{2}\right)\sin\left(\frac{k_y}{2}\right)\right]\\ &-i4E_{sx}(111)\sin\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)\\ &-i4E_{sx}(311)\sin\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)\\ &-i4E_{sx}(113)\sin\left(\frac{k_x}{2}\right)\left[\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)+\cos\left(\frac{k_z}{2}\right)\cos\left(\frac{k_y}{2}\right)\right]. \end{aligned}$$

$$\begin{aligned} H_{s1/x1}=-H^*_{s2/x2}={}&-4E_{sx}(022)\sin(k_y)\sin(k_z)\\ &+i4E_{sx}(220)[\sin(k_x)\cos(k_y)+\sin(k_x)\cos(k_z)]. \end{aligned}$$

$$\begin{aligned} H_{x1/x2}=H^*_{x2/x1}={}&4E_{xx}(111)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)\\ &+4E_{xx}(311)\left[\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)+\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_z}{2}\right)+\cos\left(\frac{k_z}{2}\right)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\right]\\ &-i4E_{xx}(111)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)\\ &+i4E_{xx}(311)\left[\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)+\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_z}{2}\right)+\sin\left(\frac{k_z}{2}\right)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\right]. \end{aligned}$$

$$\begin{aligned} H_{x1/y2}=H^*_{x2/y1}={}&H_{y1/x2}-4E_{xy}(111)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\cos\left(\frac{k_z}{2}\right)\\ &+4E_{xy}(113)\cos\left(\frac{k_z}{2}\right)\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)\\ &+4E_{xy}(311)\cos\left(\frac{k_z}{2}\right)\left[\sin\left(\frac{k_x}{2}\right)\sin\left(\frac{k_y}{2}\right)+\sin\left(\frac{k_y}{2}\right)\sin\left(\frac{k_x}{2}\right)\right]\\ &+i4E_{xy}(111)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\sin\left(\frac{k_z}{2}\right)\\ &+4E_{xy}(113)\sin\left(\frac{k_z}{2}\right)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)\\ &+i4E_{xy}(311)\sin\left(\frac{k_z}{2}\right)\left[\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right)+\cos\left(\frac{k_y}{2}\right)\cos\left(\frac{k_x}{2}\right)\right]. \end{aligned}$$

$$\begin{aligned} H_{x1/y1}=H^*_{x2/y2}={}&-4E_{xy}(220)\sin(k_x)\sin(k_y)\\ &-i4E_{xy}(022)[\cos(k_x)\sin(k_z)-\cos(k_y)\sin(k_z)]. \end{aligned}$$

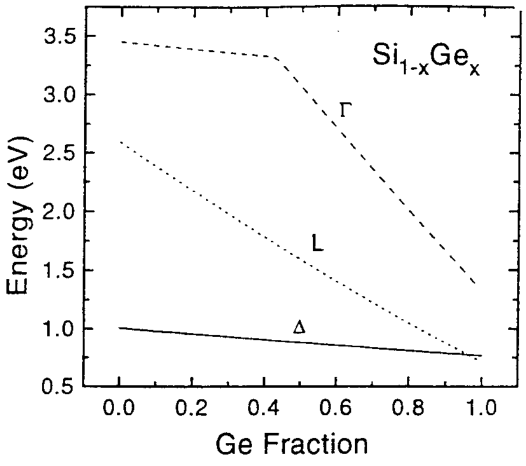


FIG. 1. Composition dependence of $Si_{1-x}Ge_x$ alloy critical points as calculated by the orthogonal LCAO method. The dependence is predominately linear. The Δ-L crossover occurs at $x=0.96$. The nonlinearity in the Γ critical point arises from a crossover to an eigensolution of differing symmetry near $x=0.4$.

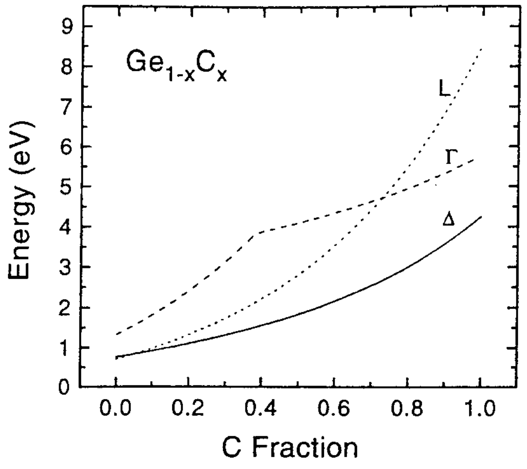


FIG. 3. Composition dependence of $Ge_{1-y}C_y$ alloy critical points as calculated by the orthogonal LCAO method. Significant nonlinearities are evident. Crossover from a L to Δ minimum is indicated for $x=0.05$.

experimental band gap (Fig. 4). Agreement is generally good, though the Δ to L crossover is predicted to occur at about 4% Si rather than the experimental value of 15%,[25] and our theory underestimates the band gap of Si. For the $Ge_{1-y}C_y$ system (Fig. 3), crossover from an L to a Δ point minimum at 5% C is predicted. The sublinear bowing departure for this system is significant—for $x<0.2$ we predict an L point dependence of 29.6 meV/% C and a Δ point dependence of 16.7 meV/% C, whereas a linear interpolation would give 78.2 and 34.9 meV/% C, respectively. For the $Si_{1-y}C_y$ system, we predict a Δ point minimum throughout the system with a composition dependence of 16.7 meV/% C for Si rich samples, rather than a linear value of 32.5 meV/% C.

Fig. 5 gives the band gap-composition dependence for a range of unstrained $Si_{1-x-y}Ge_xC_y$ alloys ($y<0.10$), which may be compared with experimental results. If measurements were obtained from relaxed alloys, a direct comparison may be made. However, in many cases, measurements were performed on thin strained alloys grown on Si substrates. In these cases, a deformation potential must be applied to remove the effects of strain on the band structure.[7,26] Because the deformation potentials for these alloys are unknown, such approaches have relied on the assumption that the deformation potentials are similar to those of Si or $Si_{1-x}Ge_x$.

Our results are compared with theoretical and experimental results of other workers in Table II. For our orthogonal LCAO calculations a dependence of 16.7 meV/% C for Δ minima and 29.6 meV/% C for L minima was predicted. Overall, a large disparity exists between the results of various workers. Some of these differences could be attributed to composition and temperature differences as well as unexpected strain behavior, but inconsistencies exist even after

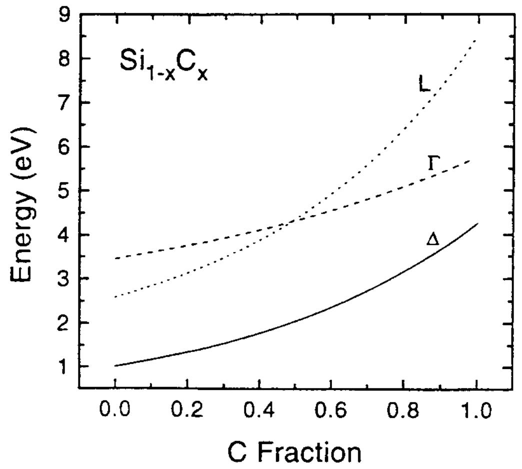


FIG. 2. Composition dependence of $Si_{1-y}C_y$ alloy critical points as calculated by the orthogonal LCAO method. The relation is nonlinear for all three critical points. A Δ minimum is indicated throughout this alloy system.

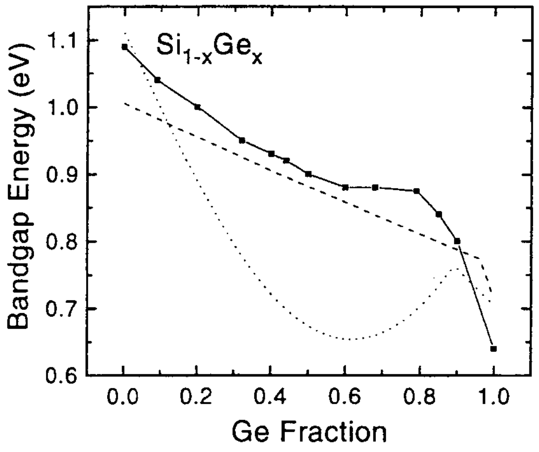


FIG. 4. Composition dependence of the $Si_{1-x}Ge_x$ band gap energy. Experimental data (see Ref. 36) is indicated by the solid line. The dashed line indicates the results of our orthogonal LCAO calculations, and the dotted line indicates the results of our nonorthogonal LCAO calculations.

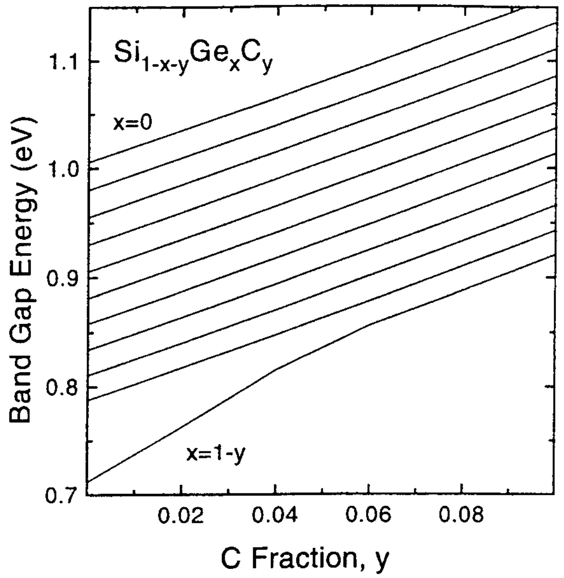


FIG. 5. Variation of the band gap energy as C is added to $Si_{1-x}Ge_x$ alloys. The top and bottom lines represent $Si_{1-y}C_y$ and $Ge_{1-y}C_y$ binary alloys, respectively. The lines represent ternary alloys of composition ratios, $x/(1-y)$, which increase in increments of 0.1 from 0 at the top to 1 at the bottom ($y \neq 1$).

considering these factors. Our LCAO results for both Δ and L minima agree well with experimental absorption results for ternary alloys, though a somewhat higher value of 63 meV/% C was experimentally found for $Ge_{1-y}C_y$ alloys. We do not obtain agreement, however, with photoluminescence (PL) and multiple quantum well (MQW) PL results for Si rich alloys, most of which indicate a variation of −7 to −19 meV/% C. Perhaps the values assumed for the deformation potential incorrectly predict the strain-free behavior. Alternatively, localized strain and alloy disorder effects, which we have not considered, could account for this decrease.

Our LCAO results for the E_1 critical point energy are in excellent agreement with spectroscopic ellipsometry results for $Si_{1-y}C_y$. The E_1 critical point energy is associated with direct transitions near the L point.[25,32] Spectroscopic ellipsometry indicates a 43 meV/% C composition dependence after strain effects are considered,[26,33,34] and our calculations predict a 40 meV/% C composition dependence. Finally, we predict a band gap energy of 2.053 eV for $Si_{0.5}C_{0.5}$, which agrees fairly well with the observed 3C-SiC band gap of 2.25 eV.[25] By comparison, the first principles calculations predict a band gap of at most 1.3 eV.[10,11,14]

By assuming Vegard's law holds for the lattice constant, and therefore alloys with a Ge:C ratio of 8:1 are lattice matched to Si, we plot the predicted band gap of the Si lattice matched $Si_{1-x-y}Ge_xC_y$ alloys in Fig. 6. We find little variation in the band gap among these alloys, however, $Si_{1-x-y}Ge_xC_y$/Si band offsets may still exist. Additionally, the band gap may be modified through strain, perhaps with greater flexibility than the $Si_{1-x}Ge_x$ system.

Fig. 7 is a band gap contour plot for the $Si_{1-x-y}Ge_xC_y$ system, giving the minimum band gaps versus composition. The uneven contour spacing indicates nonlinearities. Soref *et al.* have also found a sublinear increase in the band gap with increasing C for C lean alloys and a rapid band gap

TABLE II. C concentration dependence of the unstrained alloy band gap energy E_G. The upper section of the table shows the results of theoretical calculations; the lower section lists experimental results. The data for the $Ge_{1-y}C_y$ system assumed fully relaxed alloys. The remaining experimental data were collected on strained thin films with various C contents. The authors of these articles have estimated the deformation potentials and calculated the change in the band gap due to strain effects. By accounting for this, the band gap energy dependence on C due to alloying was computed, and is summarized here.

ΔE_G (meV/% C)	$Si_{1-x-y}Ge_xC_y$ Composition	Technique	Source
30	$x>0.96$	Orthogonal LCAO	Current study
17	$x<0.96,\ y\leq 0.10$	Orthogonal LCAO	Current study
44	$x=0$	Linear interpolation	[d]
128	$x>0.96,\ y=1-x$	Linear interpolation	[d]
48	$x<0.96,\ y=1-x$	Linear interpolation	[d]
-83	$x=0,\ y\leq 0.13$	*Ab inito* pseudo potential	[e]
−100	$x=0,\ y\leq 0.13$	*GW*	[f]
−30	$x=0,\ y\leq 0.13$	LMTO	[c]
−20	$x=0.125,\ y\leq 0.13$	LMTO	[c]
63	$x=1-y,\ y\leq 0.03$	Absorption	[g]
45	$x=0.88,\ y\leq 0.01$	Absorption	[h]
14	$x=0.08,\ y\leq 0.01$	Absorption	[i]
−19	$x=0.24, x=0.38,\ y\leq 0.011$	PL	[a]
−7	$x=0.15,\ y\leq 0.009$	MQW PL	[j]
−19	$x=0,\ y\leq 0.016$	MQW PL	[b]
67	$x=0,\ y\leq 0.014$	PL	[k]

[a] Reference 7.
[b] Reference 8.
[c] Reference 10.
[d] Reference 12.
[e] Reference 13.
[f] Reference 14.
[g] Reference 27.
[h] Reference 28.
[i] Reference 29.
[j] Reference 30.
[k] Reference 31.

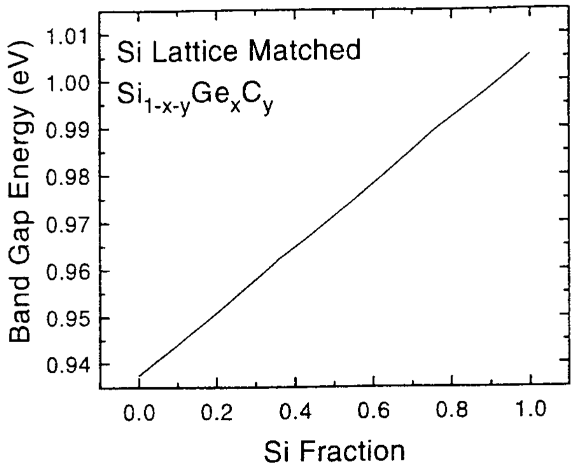


FIG. 6. Band gap energy of Si lattice matched $Si_{1-x-y}Ge_xC_y$ alloys. Vegard's law was employed to calculate the Ge:C ratio (8:1) required for lattice matching.

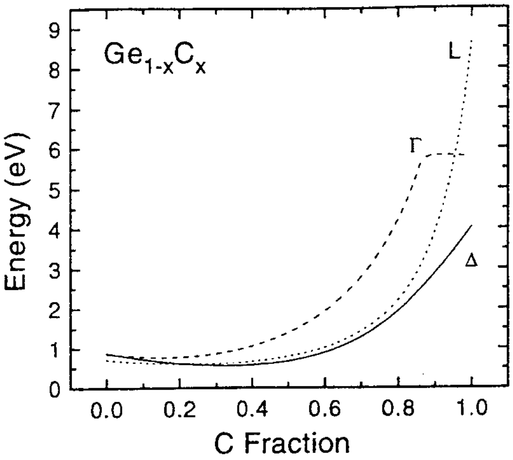


FIG. 8. Critical point energies of the $Ge_{1-y}C_y$ alloy system as calculated by the nonorthogonal LCAO method. Critical point energies are predicted to initially decrease as C is added to Ge, then rise rapidly for C rich alloys.

increase for C rich alloys[35] in a plot obtained by combining LMTO theoretical results for Si rich alloys[10] with experimental results for Ge rich alloys.[9] Because the LMTO results were included, he did find a band gap decrease as C content increased for some Si rich alloys.

The non-orthogonal calculations produced more pronounced nonlinearities. For the Ge rich $Si_{1-x-y}Ge_xC_y$ alloys and $Ge_{1-y}C_y$ alloys (Fig. 8) a decrease in the band gap with increasing C is predicted for some alloys with Δ minimums. $Si_{1-y}C_y$ alloys (Fig. 9) show a decrease only in the L minimum, which is never the conduction band minimum for this system, and thus the band gap of the $Si_{1-y}C_y$ system is indicated by this calculation to be monotonically increasing. Similarly the theoretical $Si_{0.5}C_{0.5}$ band gap energy of 1.608 eV does not agree as well as the orthogonal method. As shown in Fig. 4, the bowing departure that is observed experimentally for the $Si_{1-x}Ge_x$ system is overestimated, indicating a non-monotonic band gap variation that is not experimentally observed.

As expected from the increased number of parameters, the nonorthogonal method produced superior results at the end points but appear to deviate at the midpoints of the interpolation. We consider the orthogonal method to be a better representation of the physical alloy band structure due to its improved match to the experimental band gaps in the $Si_{1-x}Ge_x$ system, and to 3C-SiC. The nonorthogonal method may ultimately provide a superior description of the alloy band structure if an improved parameter set could be found. In principle, experimental results for critical point energies of alloys could be included in the fitting algorithm used to determine the parameter set, and the non-orthogonal LCAO results could then be used to interpolate the band structure

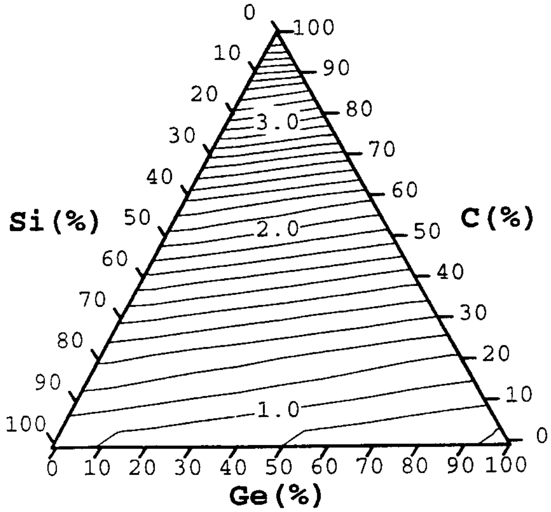


FIG. 7. Contour plot of band gap energies of $Si_{1-x-y}Ge_xC_y$ alloys as calculated by the orthogonal LCAO method. Units for contour labels are eV; each contour line represents a step of 0.1 eV. Uneven contour spacing illustrates nonlinearities, particularly evident for C rich alloys. Band gap energy variations are monotonic with composition.

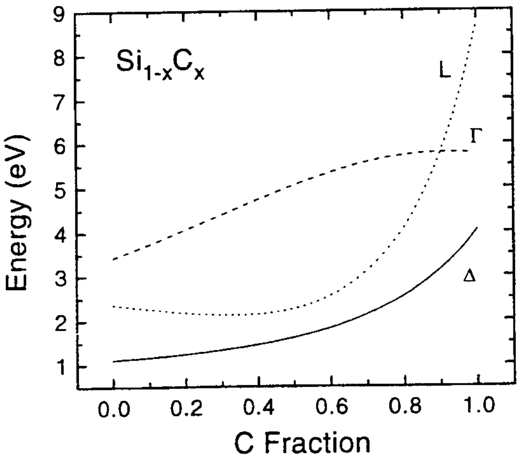


FIG. 9. Critical point energies of the $Si_{1-y}C_y$ system as calculated by the nonorthogonal LCAO method. The L point energy initially decreases as C is added to Si, but the conduction band minimum at Δ increases slightly.

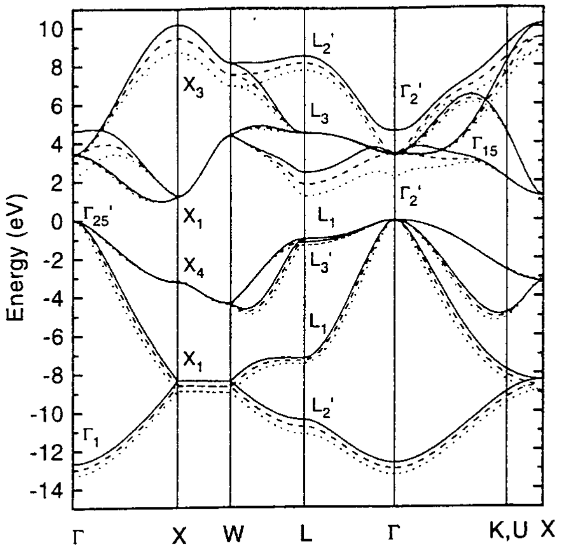


FIG. 10. Band structure of $Si_{1-x-y}Ge_xC_y$ alloys lattice matched to Si calculated by the orthogonal LCAO method. Alloys consisted of 90% Si (solid lines), 50% Si (dashed lines), and 10% Si (dotted lines). The remaining alloy components were Ge and C in an 8:1 ratio. The valance band structure is virtually unchanged. The conduction bands have a similar structure throughout, with the most noticeable changes being a lowering with decreasing Si in energy of the minimums at L and Γ, and a change in symmetry of the conduction band minimum at Γ.

for other wave vectors and other alloys. Presently, experimental data is sparse for alloys containing C.

Complete band structures were calculated for selected alloys using the orthogonal LCAO method. Fig. 10 gives the band structure of a series of lattice matched $Si_{1-x-y}Ge_xC_y$ alloys. Figs. 10–13 give the band structure of several C lean binary and ternary alloys, illustrating changes due to the addition of C. Fig. 14 shows the band structure of a 3C-SiC lattice matched (by Vegard's law) $Ge_{1-y}C_y$ alloy. The resulting band structures vary gradually with composition. In general, alloys with one dominant element had a band structure characteristic of that element. A Δ minimum is expected for most alloys, except those with at least 95% Ge. None of the alloys was predicted to have a direct band gap for any composition.

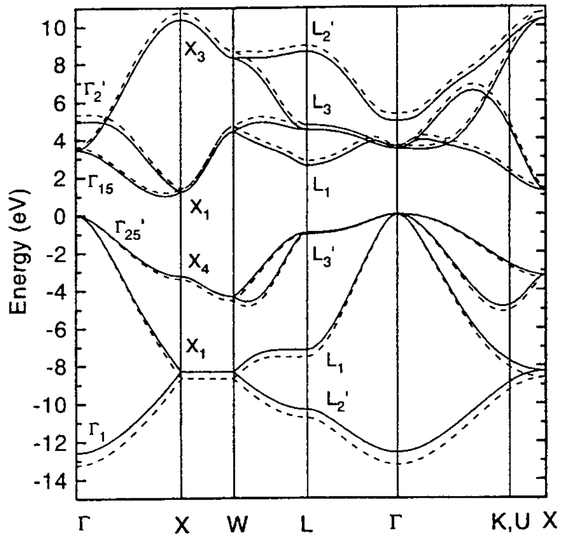


FIG. 11. Band structure of Si (solid lines), and $Si_{0.90}C_{0.10}$ (dashed lines) as calculated by the orthogonal LCAO method. There are no significant changes in the band structure expected other than an increase in the band gap energy with C.

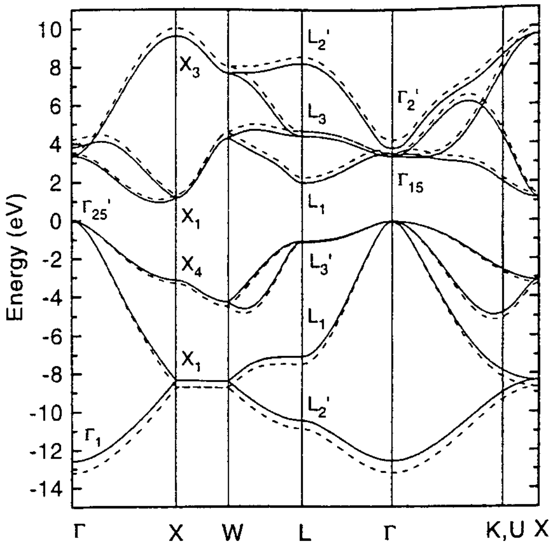


FIG. 12. Band structure of $Si_{0.70}Ge_{0.30}$ (solid lines) and $Si_{0.63}Ge_{0.27}C_{0.10}$ (dashed lines) as calculated by the orthogonal LCAO method.

IV. CONCLUSION

In conclusion, we have calculated the band structure for the $Si_{1-x-y}Ge_xC_y$ alloy system by the empirical LCAO method. An indirect Δ minimum was indicated for most alloys, and the critical point energies were found to have a

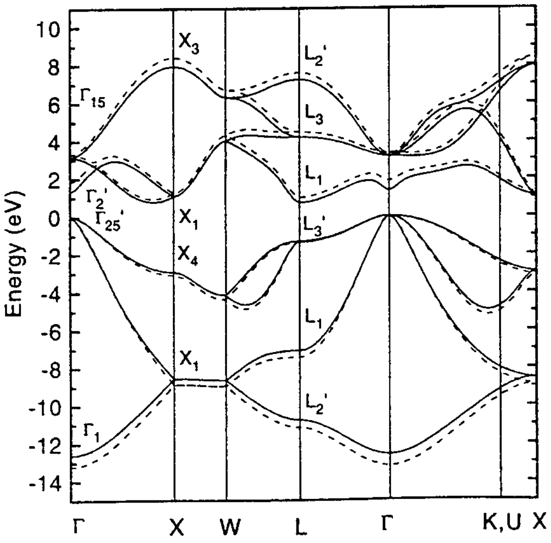


FIG. 13. Band structure of Ge (solid lines) and $Ge_{0.90}C_{0.10}$ (dashed lines) as calculated by the orthogonal LCAO method.

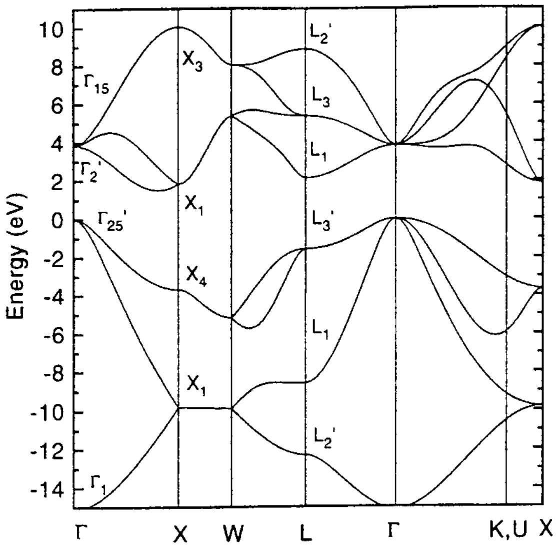


FIG. 14. Band structure of $Ge_{0.62}C_{0.38}$ as calculated by the orthogonal LCAO method. By Vegard's law, this alloy is lattice matched to 3C-SiC.

monotonic but non-linear composition dependence. Our goal was to show results of this band structure approach for comparison with other published calculations.

ACKNOWLEDGMENTS

The authors are grateful to Dr. M. Berding, Dr. P. Christie, and Dr. R. Soref for useful discussions and review of this manuscript. This work was supported by ONR Grant No. N00014-93-1-0393 and DARPA Contract No. F49620-96-C-0006.

[1] J. C. Bean, Proc. IEEE **80**, 571 (1992).

[2] R. A. Soref, Proc. IEEE **81**, 1687 (1993).

[3] H. J. Osten, J. Cryst. Growth **157**, 268 (1995).

[4] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, Appl. Phys. Lett. **60**, 3033 (1992).

[5] J. W. Strane, H. J. Stein, S. R. Lee, B. L. Doyle, S. T. Picraux, and J. W. Mayer, Appl. Phys. Lett. **63**, 2786 (1993).

[6] R. I. Scace and G. A. Slack, J. Chem. Phys. **30**, 1551 (1959).

[7] A. St. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewalt, Appl. Phys. Lett. **67**, 3915 (1995).

[8] K. Brunner, K. Eberl, and W. Winter, Phys. Rev. Lett. **76**, 303 (1996).

[9] J. Kolodzey, P. R. Berger, B. A. Orner, D. Hits, F. Chen, A. Khan, X. Shao, M. Waite, S. I. Shah, C. P. Swann, and K. M. Unruh, J. Cryst. Growth **157**, 386 (1995).

[10] J. Xie, K. Zhang, and X. Xie, J. Appl. Phys. **77**, 3868 (1995).

[11] M. A. Berding, A. Sher, and M. van Schilfgaarde, Phys. Rev. B (in press).

[12] R. A. Soref, J. Appl. Phys. **70**, 2470 (1991).

[13] A. A. Demkov and O. F. Sankey, Phys. Rev. B **48**, 2207 (1993).

[14] J. Gryko and O. F. Sankey, Phys. Rev. B **51**, 7295 (1995).

[15] K. E. Newman and J. D. Dow, Phys. Rev. B **30**, 1929 (1984).

[16] G. Theodorou, N. D. Vlachos, and C. Tserbak, J. Appl. Phys. **76**, 5294 (1994).

[17] P. Soven, Phys. Rev. **156**, 809 (1967).

[18] D. Stroud and H. Ehreneich, Phys. Rev. B **2**, 3197 (1970).

[19] S. Krishnamurthy, A. Sher, and A. B. Chen, Phys. Rev. B **33**, 1026 (1986).

[20] D. A. Papaconstantopoulos, *Handbook of the Band Structure of Elemental Solids* (Plenum, New York, 1986).

[21] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

[22] L. F. Mattheiss, Phys. Rev. B **2**, 3918 (1970).

[23] P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids **44**, 365 (1981).

[24] W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980).

[25] *Landolt-Bornstein: Numerical Data and Functional Relationships in Science and Technology, Group III*, edited by O. Madelung (Springer, Berlin, 1982), Vol. 22.

[26] S. Zollner, J. Appl. Phys. **78**, 5209 (1995).

[27] B. A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R. G. Wilson, P. R. Berger, and J. Kolodzey, J. Electron. Mater. **25**, 297 (1996).

[28] B. A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J. W. Mayer, and J. Spear, Appl. Phys. Lett. **69**, 2557 (1996).

[29] B. A. Orner, D. Hits, J. Kolodzey, F. J. Guarin, A. R. Powell, and S. S. Iyer, J. Appl. Phys. **79**, 8656 (1996).

[30] P. Boucaud, C. Guedj, D. Bouchier, F. H. Julien, J. M. Lourtioz, S. Bodnar, J. L. Regolini, and E. Finkman, J. Cryst. Growth **157**, (1995).

[31] H. Lee, S. R. Kurtz, J. A. Floro, J. Strane, C. H. Seager, S. R. Lee, E. D. Jones, J. F. Nelson, T. Mayer, and S. T. Picraux, Jpn. J. Appl. Phys. 1 **34**, L1340 (1995).

[32] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors* (Springer, Berlin, 1989).

[33] W. Kissinger, M. Weidner, H. Osten, and M. Eichler, Appl. Phys. Lett. **65**, 3356 (1994).

[34] W. Kissinger, H. Osten, M. Weidner, and M. Eichler, Appl. Phys. Lett. **79**, 3016 (1996).

[35] R. A. Soref, Z. Atzman, F. Shaapur, M. Robinson, and R. Westhoff, Opt. Lett. **21**, 345 (1996).

[36] R. Braunstein, A. R. Moore, and F. Herman, Phys. Rev. **109**, 695 (1958).

Journal of Electronic Materials, Vol. 26, No. 12, 1997

Regular Issue Paper

Electrical Properties of $Si_{1-x-y}Ge_xC_y$ and $Ge_{1-y}C_y$ Alloys

F. CHEN,[*] R.T. TRÖGER,[*] K. ROE,[*] M.D. DASHELL,[*] R. JONCZYK,[*]
D.S. HOLMES,[†] R.G. WILSON,[‡] and J. KOLODZEY[*]

[*]Electrical Engineering Department, 140 Evans Hall, University of Delaware, Newark, DE 19716

[†]Lake Shore Cryotronics, Inc., 64 East Walnut Street, Westerville, OH 43081-2399

[‡]Hughes Research Lab, 3011 Malibu Canyon Road, Malibu, CA 90265

The effects of carbon on the structural and electrical properties of $Si_{1-x-y}Ge_xC_y$ and $Ge_{1-y}C_y$ alloys grown by the molecular beam epitaxy have been examined by Hall effect measurements, current-voltage measurements, x-ray diffraction, and atomic force microscopy. Hall effect measurements showed that the addition of carbon increased the hole mobility in GeC compared to pure Ge, we attributed this increase to improved crystalline quality and reduced surface roughness. Current-voltage characteristics of SiGeC/Si and GeC/Si heterojunction diodes showed that with increasing carbon, the reverse leakage current decreased and the forward turn-on voltage increased, attributed to the increase in bandgap energy and reduction of intrinsic carrier concentration n_i.

Key words: Carbon, electrical properties, $Ge_{1-y}C_y$, germanium, group IV alloys, Hall mobility, I-V characteristics, molecular beam epitaxy (MBE), $Si_{1-x-y}Ge_xC_y$

INTRODUCTION

Recently, the growth and characterization of the Group IV elements (C, Si Ge) have been attracting attention for use in heterojunction (HJ) devices compatible with Si technology.[1] Although the equilibrium solubility of C in Si and Ge is low, $Si_{1-x-y}Ge_xC_y$ and $Ge_{1-y}C_y$ alloys can be synthesized at relatively low growth temperatures[2] by non-equilibrium growth methods such as molecular beam epitaxy (MBE). The addition of C to SiGe and Ge affects the lattice constants, strain compensation, carrier transport, bandgap energy, and the energy band offsets. Although the structure and optical properties of $Si_{1-x-y}Ge_xC_y$ and $Ge_{1-y}C_y$ alloys have been reported,[3,4] the electrical properties are not yet well understood.

This paper focuses on the electrical properties of p-type Ge-rich SiGeC and GeC alloys grown on n-type Si-substrates. Hall effect measurements have been used to study the effects of carbon on the charge transport, and current-voltage (I-V) characteristics of SiGeC/Si and GeC/Si HJ diodes have been used to study the reverse leakage current and forward turn-on voltage.

EXPERIMENTS

The SiGeC and GeC epilayer were grown by molecular beam epitaxy in an EPI620 system as described elsewhere.[5] The Si was evaporated from a solid thermal source using a pyrolytic graphite crucible. The Ge was evaporated from a solid thermal source using a pyrolytic boron nitride crucible. The C beam was produced by sublimation from a pyrolytic graphite filament. Elemental boron was evaporated from a high temperature effusion cell with a graphite crucible inserted into a tungsten jacket. N-type Si (100) substrates were prepared by degreasing, etching, and an HF dip. Growth occurred at substrate temperatures between 400 and 550°C. The lattice constants of our epilayers perpendicular to the surface $a_\perp$ were determined from the symmetrical (004) reflection by x-ray diffractometry using θ-2θ-scan-

(Received July 29,1996; accepted June 3, 1997)

Table I. Measured Properties of SiGeC and GeC Alloys Including Compositions and Thickness

	X-Ray Lattice ($a_\perp$) Constant (Å)	Thickness (µm)	SIMS B Concentration (cm^{-3})	X-Ray Composition	XPS Composition
SGC83 Ge	5.66356	0.59	3×10^{18}	C:0%	
SGC80 GeC	5.66258	0.56	3×10^{18}	C:0.27%	
SGC84 GeC	5.66239	0.59	5×10^{18}	C:0.33%	
SGC73 SiGeC	5.64992	0.18	1×10^{19}		Si:7%; C:2%; Ge:91%

Note: The composition has been deduced from linear extrapolations between the lattice constant in GeC and Ge, calibrated from higher C concentrations measured by resonant RBS. The composition results are consistent with growth conditions.

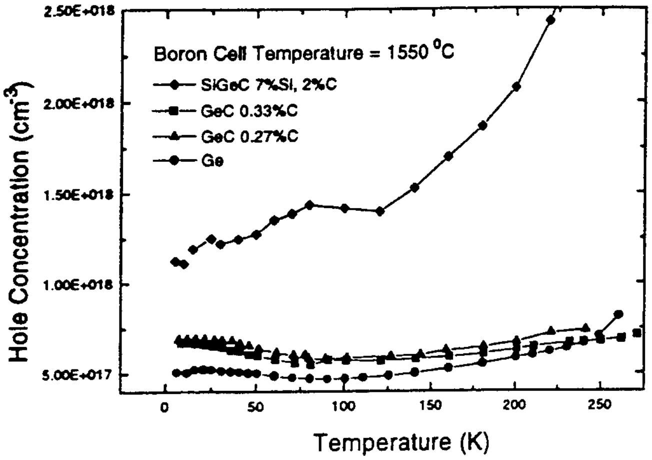


Fig. 1. Hole concentration vs temperature measured by Hall effect for four Ge-rich crystal alloys at temperature range between 10 and 250K with the boron cell temperature at 1550°C. The SiGeC sample had a slightly higher hole concentration than the other samples.

ning. The chemical composition for a set of Ge-rich GeC alloys was deduced from linear interpolations between the lattice constants in GeC and Ge, calibrated from higher C concentrations measured by resonant Rutherford backscattering spectroscopy (RBS). The chemical composition for a Ge-rich SiGeC alloy was measured by x-ray photoelectron spectroscopy (XPS) and by secondary ion mass spectrometry (SIMS). The results are presented in Table I. Secondary ion mass spectrometry indicated that the boron distribution was uniform throughout the doped layers for all our samples.

Hall effect measurements were carried out by using a standard van der Pauw structure with In metal contacts under a magnetic field of 0.4 T. The measurement accuracy was about 5%. The alloy layers were electrically isolated from the substrate using a reverse biased p-n junction. Carrier transport occurred primarily in the p-type epitaxial alloy at sufficiently low biases, currents and low temperature.

Standard optical lithography and wet chemical etching were used to form mesas for the SiGeC/Si and GeC /Si HJ diodes. The etchant H_3PO_4:H_2O_2:H_2O (1:6:3) was used to etch the SiGeC and GeC epilayers. Ti/Au metal contacts of thickness 300Å/1500Å, respectively, were evaporated using an electron beam, and patterned by the standard lift-off method. The Ti/Au contacts were measured to have ohmic characteristics over the full measurement range.

RESULTS AND DISCUSSIONS

Carbon Effect on Hole Mobility

As shown in Table I, SIMS results show that the B concentration increases slightly with C content for a constant B flux, possibly due to an increased B sticking coefficient caused by the C during growth. The SiGeC alloy has a higher doping concentration corresponding to the higher B incorporation rate than either the Ge or the GeC alloys. We conclude that both Si and C increase the B incorporation compared to pure Ge. Figure 1 shows the hole concentration vs temperature measured by Hall effect. The Ge and GeC samples have hole concentration less than the SIMS B concentration. This observation could be due to low B activation or to calibration errors in the SIMS.

Figure 2a shows the temperature and concentration dependence of Hall mobility in Ge-rich alloys with identical B flux. For the Ge and GeC samples, the hole mobility increased with increasing C. Since the B acceptor concentration is the same or higher with increasing C content, it is unlikely that reduced impurity scattering can explain the increase of mobility with C. Since the samples are of the same thickness, the increase of µ with C fraction appears to be intrinsic. The maximum mobility for these three samples was at temperature T ≈ 160K. The dependence of Hall mobility on hole concentration is plotted in Fig. 2b. The sample with the higher C concentration had higher mobility. For the B doped pure Ge layer, the hole mobility is close to the value for bulk Ge.

We speculate that the aperiodic atomic potentials of substitutional[6] and interstitional[7] carbon in Ge may increase the scattering rates, and reducing the carrier mobility. On the other hand, however, there are several possibilities for C to increase the mobility. These include the improvement in crystalline quality,[8] the reduction of effective mass for carriers, and the reduction of surface roughness scattering.

From the x-ray diffraction measurements, the full width at half maximum (FWHM) of the diffraction angle θ for the (004) diffraction line was used as an indication of crystalline quality. The values of FWHM for our samples are reported in Table II. TEM studies show that our epilayers are dislocated primarily by stacking faults and twins. For the Ge and GeC samples, adding C to Ge improved the crystalline quality by reducing the FWHM. One possible reason for this improvement is that the strain is reduced by the C.[9] The improvement of the crystalline quality correlates well to the observed increase in mobility.[10]

The surface roughness was determined by atomic force microscopy (AFM) surface scanning. Figure 3 shows the two AFM images for pure Ge and GeC samples. As we see from Fig. 3, C greatly smoothed out the surface roughness. Perhaps this was resulted by partial strain compensation and defect formation suppression. Such roughness can be modeled to act as a spatial variation of the position of the band discontinuity at the surface.[11] The mobility could be limited by surface roughness scattering.[12] The value of average surface roughness for some of our samples are listed in Table II.

For the particular case of our SiGeC alloy, the problem of decreasing crystalline quality and increasing surface roughness observed by adding Si to GeC is attributed to our thermal Si source which uses a graphite crucible. It may be that some C is co-evaporated with the Si. Due to our system limitation, our SGC73 (SiGeC) has the largest FWHM, the highest surface roughness, the biggest compositional difference, and the smallest mobility.

Influence of Carbon on I-V Characteristics of SiGeC/Si and GeC/Si Heterojunction Diodes

A major drawback of binary SiGe/Si for high performance electronic devices is the relatively large lattice mismatch between Si and Ge (~4%). If the layer thickness exceeds a critical value, mismatch leads to strain-relieving misfit dislocations which may introduce states in the bandgap due to their strain fields and dangling bonds. Dislocations can introduce generation-recombination centers which can increase reverse leakage current [13] and decrease breakdown voltage.

C affects the electrical and structural properties of GeC and SiGeC,[14] and may affect the diode I-V characteristics. Possible mechanisms for C to affect device electrical properties include the bandgap, band offsets,[15] and the suppression of B diffusion.[16] The possible effects of C on structural properties include the mismatch between the hetero-interface, the crystalline quality, and localized tensile sites that act as pinning centers to block the propagation of dislocations.[6]

Experimentally, most HJs have the following general formulations for the current density[17]

$$J = J_{00}\exp(-\Delta E/kT)[\exp(qV/\eta kT)-1]$$

where J_{00} is the saturation reverse current, η is the ideality factor, and ΔE is a measured activation en-

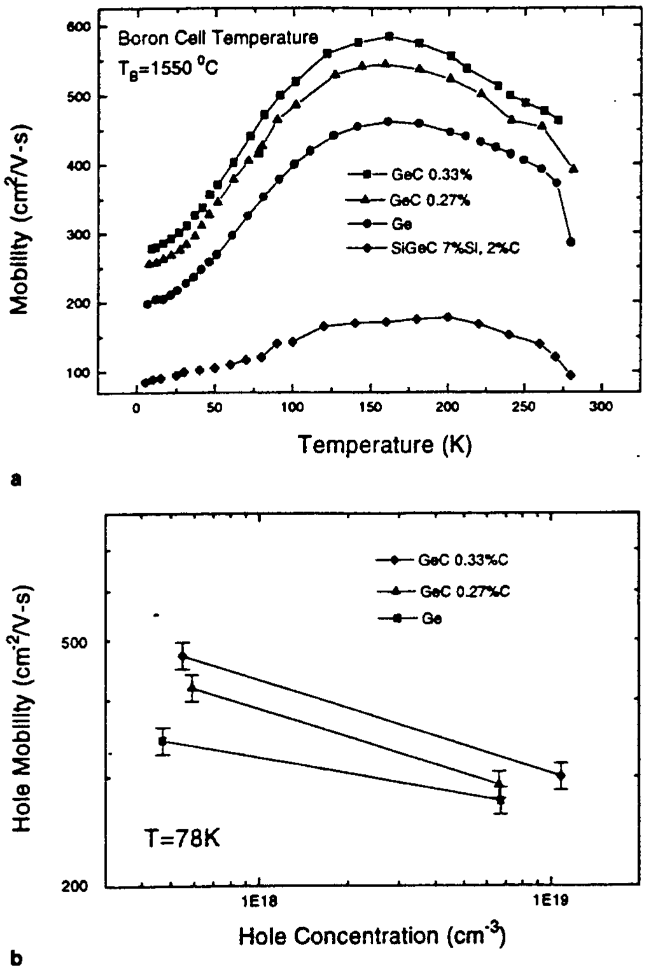
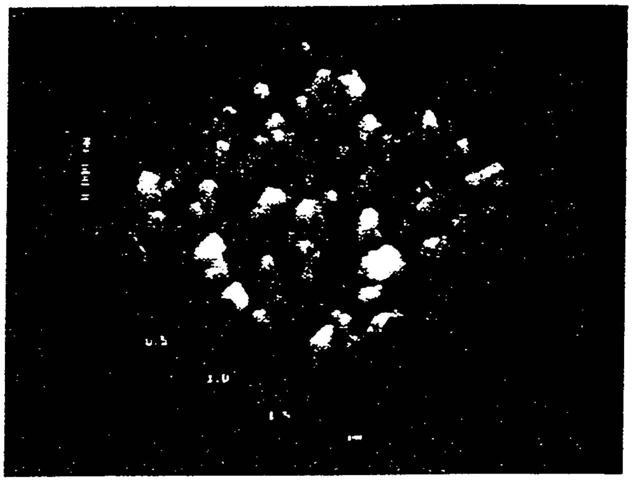


Fig. 2. (a) Temperature dependence of Hall mobility for four Ge-rich crystal alloys in the temperature range between 10 and 250K. The hole mobility increased with increasing C for GeC alloys. SiGeC alloy had the lowest mobility. The maximum mobility for these four alloys occurred near T = 160K. (b) Hole concentration dependence of Hall mobility at T = 77K for the same alloys. The sample with the highest C concentration consistently has the highest mobility.

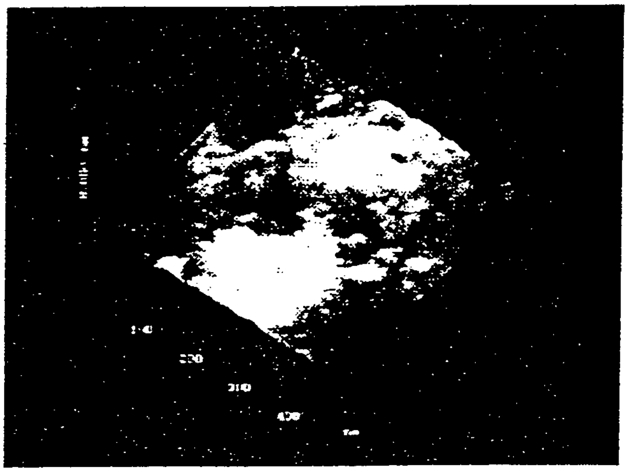
Table II. FWHM of XRD (004) Peak Intensity for SiGeC and GeC Alloys

	FWHM (θ)	Mean Surface Roughness (nm)	Activation Energy from I-V (eV)
$Si_{0.07}Ge_{0.91}C_{0.02}$	0.497	15.21	0.38
Ge	0.310	0.473	0.40
$Ge_{0.9973}C_{0.0027}$	0.290	0.347	0.51
$Ge_{0.9967}C_{0.0033}$	0.139	0.308	0.62

Note: Mean surface roughness measured by AFM. By increasing the C in the Ge, the FWHM decreased, the mean roughness decreased, and the crystalline quality was improved. Activation energy for the zero bias extrapolation of the lnJ vs V curves was obtained from an Arrhenius plot for different temperature. The forward saturation current density J_{00} exhibits an exponential dependence on temperature implying thermally activated behavior. For GeC, adding carbon increases the activation energy.



Ge



GeC with 0.27% C

Fig. 3. Surface roughness image measured by AFM. Adding C smoothed out the surface roughness due to partial strain compensation and defect formation suppression.

ergy for the zero bias extrapolation of the forward current. All our diodes exhibit strong rectifying behavior with ideality factor $\eta \approx 1$. From the values of η, and the ΔE obtained from an Arrhenius plot for different temperatures, we infer that the principal current mechanism is caused by carriers injection over a barrier.[18] Due to large valence band offsets,[15] we conclude that the electrons inject from the wider gap n-Si to the p^+-GeC or SiGeC layer. The activation energy ΔE for our diodes is shown in Table II.

As shown in Fig. 4, the reverse leakage current decreases with the addition of C. Small amounts of C (0.3%) reduce the reverse leakage current by a factor of three compared to pure Ge. Because of heavy doping, the depletion layer occurs mainly on the n-side (Si) of the junction. The estimated maximum depletion layer thickness on the p-side (10^{18} cm^{-3}) of the junction is about 1.5 nm with the n-type 1–10Ω–cm substrate. The depletion region of the Si side is approximate 150 nm, which is similar for all diodes.

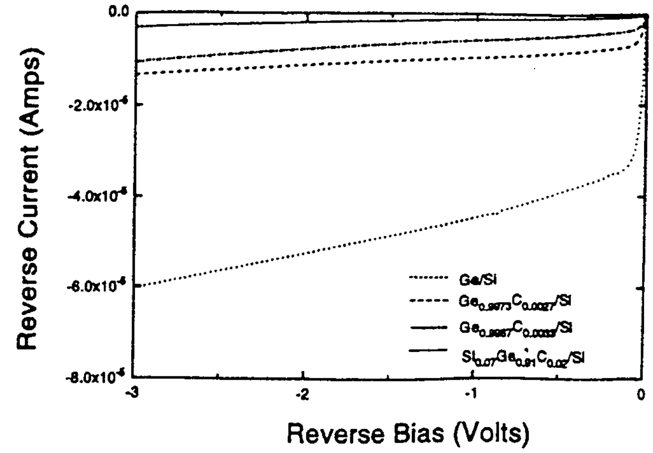


Fig. 4. The reverse leakage current of four Ge-rich crystal alloys. The device area was 2.2×10^{-3} cm^2 for all devices. Leakage current decreased with increasing C.

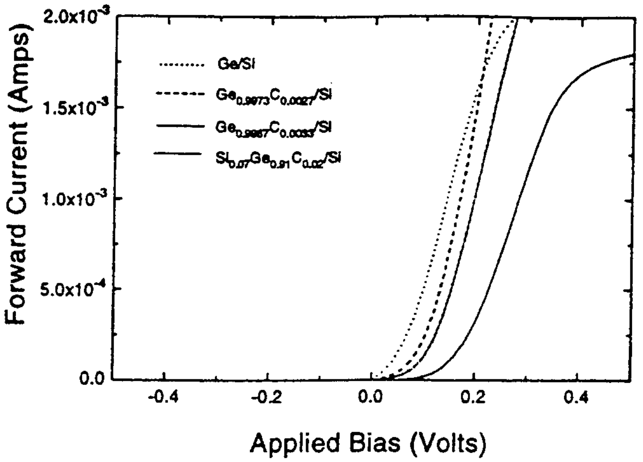


Fig. 5. Forward I-V characteristics for four Ge-rich crystal alloys. Device area was 2.2×10^{-3}cm^2. C increased the turn-on voltage by the amount of 0.3V/%C.

Therefore, depletion layer contributions, such as generation current, are the same for all diodes and cannot explain the observed differences with composition. The SiGeC sample exhibits the smallest reverse leakage current, however, it has the poorest crystalline quality and the largest surface roughness. Thus, the origin of the reverse leakage current difference cannot be attributed to the crystal quality as given by the x-ray full width at half maximum (FWHM). Instead, the observed changes in reverse current are possibly related to differences in n_i which depends on bandgap and to interface effects such as heterointerface defects. Figure 5 shows the forward I-V curves for these four samples. C increases the forward turn-on voltage. High turn-on voltages usually correspond to larger bandgaps, suggesting that adding C to Ge will increase the bandgap.

In order to understand the mechanism for the I-V characteristic with adding C, we have performed the optical absorption Fourier transform infrared (FTIR) measurements to find the bandgap for some samples. The FTIR results are plotted in Fig. 6, showing a blue shift of the absorption edge of Ge-rich alloys with increasing C, consistent with our electrical measurements. It may be that C decreases the bandgap of Si-rich alloys. The difference of the bandgap may be the dominate reason for the decreasing reverse leakage current with C. Higher bandgaps might produce smaller leakage current and higher forward turn-on voltage. We also note that the activation values increase with increasing bandgap for GeC samples, but not for the SiGeC sample which has the smallest activation but largest bandgap. The differences in activation between GeC and the SiGeC alloys are possibly related to the difference of the conduction band offsets between heterojunction with composition. We observed that C incorporation into SiGe and Ge yield better I-V characteristics with smaller leakage current and higher turn-on voltage.

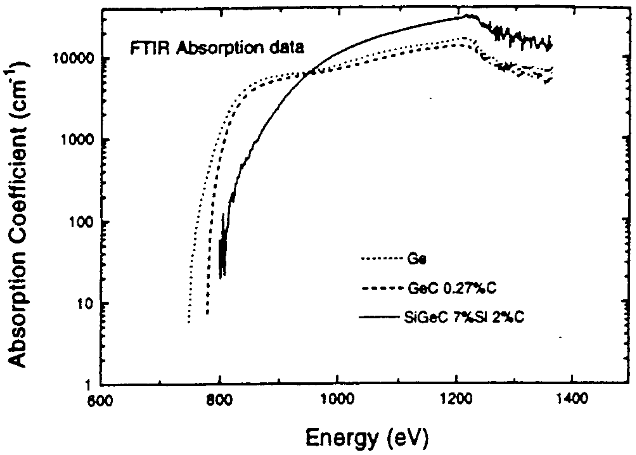


Fig. 6. FTIR optical absorption results for three of our samples. Adding C in Ge produced a blue shift of absorption edge. SiGeC sample had the highest bandgap related to the best I-V characteristics.

CONCLUSIONS

In this paper, the electrical properties of SiGeC and GeC alloys have been examined. Adding C to Ge increased the carrier mobility for thin epilayer. C in SiGe and Ge can decrease the average lattice constant and the strain, affect the crystalline quality, influence the surface roughness, and change the bandgap and band offsets. This resulted in a better I-V characteristics with smaller leakage current and higher turn-on voltage. Our electrical study indicated that by careful control of the growth conditions and C percentages, high-quality devices can be fabricated. SiGeC and GeC alloys may open up an exciting new region for HJ devices with Si technology compatible applications.

ACKNOWLEDGMENT

The authors gratefully acknowledge Mr. D. Guerin and Dr. S. Ismat Shah for XPS measurements, and Mr. A. Khan and Dr. P. Berger for wet chemical etching, and to Dr. M. Barteau for the AFM. This work was supported by grants from AFSOR (F49620-95-1-0135), ARO (DAAH04-95-10625), DARPA (F49620-96-C-0006), and the ONR (N00014-93-1-0393).

REFERENCES

1. R.A. Soref, *Proc. IEEE* 81, 1687 (1993).
2. A.R. Powell, K. Eberl, B.A. Ek and S.S. Iyer, *J. Cryst. Growth* 127, 425 (1993).
3. J. Kolodzey, P.R. Berger, B.A. Orner, D. Hits, F. Chen, A. Khan, X. Shao, M.M. Waite, S. Ismat Shah, C.P. Swann and K.M. Unruh, *J. Cryst. Growth* 157, 386 (1995).
4. B.A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R.G. Wilson, P.R. Berger and J. Kolodzey, *J. Electron. Mater.* 25, 297 (1996).
5. J. Kolodzey, P.A. O'Neil, S. Zhang, B.A. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite and S. Ismat Shah, *App. Phys. Lett.* 67, 1865 (1995).
6. C. Guedj, D. Bouchier, P. Boucaud, G. Hincelin, X. Portier, A.L. Hoir, S. Bodnar and J.L. Regolini, *Mater. Sci. and Eng. B* 36, 286 (1996).
7. G. Daries and R.C. Newman, *Handbook of Semiconductors*, vol. 3, ed. T.S. Moss (Amsterdam: Elsevier Science, 1994).
8. A. Fukami, K. Shoji, T. Nagano and C.Y. Yang, *Appl. Phys. Lett.* 57, 2345 (1990).
9. S. Lombardo, F. Priolo, S.U. Campisano and S. Lagomarsino, *Appl. Phys. Lett.* 62, 2335 (1993).
10. F. Zhang, N. Zhang, E. Xie, B. Yang, Y. Yang, G. Chen and X. Jiang, *Thin Solid Films* 216, 279 (1992).
11. T. Ando, *J. Phys. Soc. Jpn.* 51, 3900 (1982).
12. C. Weisbuch and B. Vinter, *Quantum Semiconductor Structures*, (San Diego: Academic Press, Inc., 1991).
13. E.A. Beam III, H. Temkin and S. Mahajam, *Semi. Sci. and Tech.* 7, 229 (1992).
14. K. Erbel, S.S. Iyer, S. Zollner, J.C. Tsang and F.K. LeGoues, *Appl. Phys. Lett.* 60, 3033 (1992).
15. F. Chen, M.M. Waite, S. Ismat Shah, B.A. Orner, S.S. Iyer and J. Kolodzey, *App. Surf. Sci.* 104–105, 615 (1996).
16. S. Nishikawa, A. Tanaka and T. Yamaji, *Appl. Phys. Lett.* 60, 2770 (1992).
17. A.L. Fahrenbruch and R.H. Bube, *Fundamentals of Solar Cells*, (New York: Academic Press, 1983).
18. F. Chen, B.A. Orner, D. Guerin, A. Khan, P.R. Berger, S. Ismat Shah and J. Kolodzey, *IEEE Elec. Dev. Lett.* 17, 589 (1996).

$Ge_{1-x}C_x$/Si Heterojunction Photodiodes

Xiaoping Shao, S. L. Rommel, B. A. Orner, H. Feng, M. Dashiell, J. Kolodzey, and Paul R. Berger
Department of Electrical Engineering
University of Delaware
140 Evans Hall, Newark, DE 19716

ABSTRACT

We report on the fabrication and characterization of a photodiode made from a heterojunction of epitaxial p-type $Ge_{1-x}C_x$ on an n-type Si substrate. Epitaxial $Ge_{1-x}C_x$ layers with carbon percentages of 0.2, 0.8, 1.4, and 2% were grown on (100) Si substrates by solid source molecular beam epitaxy. The p-GeC/n-Si junction exhibits diode rectification with low reverse saturation current ($\leq$ 89 pA/μm^2 at -1 volt) and high reverse breakdown voltage in excess of -40 volts. Despite the large number of dislocations and defects at the heterojunction, photoresponsivity was observed from the p-$Ge_{1-x}C_x$/n-Si diodes using laser excitation at a wavelength of 1.3 μm. External quantum efficiency was measured between 1.2 and 2.3%.

Keywords: Group IV alloys, germanium carbon, heterojunction, photodiode, photoresponsivity

1. INTRODUCTION

SiGeC/Si heterostructures offer potential applications for Si-based electronic as well as optoelectronic integrated circuits in the wavelength region of 1.3 μm suitable for fiber-optic communications.[1,2] In the past decade, most investigations have focused upon SiGe alloys on Si substrates. In order to obtain efficient photoresponse in the 1.3 μm wavelength region, the Ge concentration should be greater than 50%. However, due to the 4% lattice mismatch between Si and Ge, the critical thickness for a strained pseudomorphic SiGe epilayer[2] with this composition is limited to less than 100 Å. Therefore, there is a serious constraint in design of SiGe heterojunction devices. Incorporation of C into the SiGe system should reduce the compressive strain between the SiGe epilayer and the Si substrate below, and therefore makes it possible to control the strain within the ternary alloy SiGeC system about the lattice matching condition for Si substrates while allowing adjustable bandgaps.[3-5] According to Vegard's law, lattice matching to Si can be achieved, by use of a Ge to C ratio of 8.2 to 1, and simultaneously this can be diluted with zero to 100% Si. Our research has concentrated on GeC alloys. There is considerable difficulty in obtaining large fractions of C in Ge due its low solubility of 10^8 atoms/cm^3 at the melting point of Ge.[6] GeC is thermodynamically unstable in solid state form,[7] decomposing into its segregated components, under zero pressure. But, non-equilibrium growth techniques, such as molecular beam epitaxy (MBE) or chemical vapor deposition (CVD), have been successful in synthesizing $Ge_{1-x}C_x$ alloys with small percentages of carbon, under 3% by MBE[8] and 5% by CVD.[9] Research on the $Ge_{1-x}C_x$ system has centered mostly upon investigation of material properties. Our investigations have begun to explore processing issues[10] and now device performance, using these new Group IV alloys. In this paper, we report on the fabrication and characterization of a p-$Ge_{1-x}C_x$/n-Si heterojunction photodiode grown by MBE with 0.2 - 2% C. Our results demonstrate rectification and photoresponsivity for this p-GeC/n-Si photodiode.

2. EXPERIMENTS

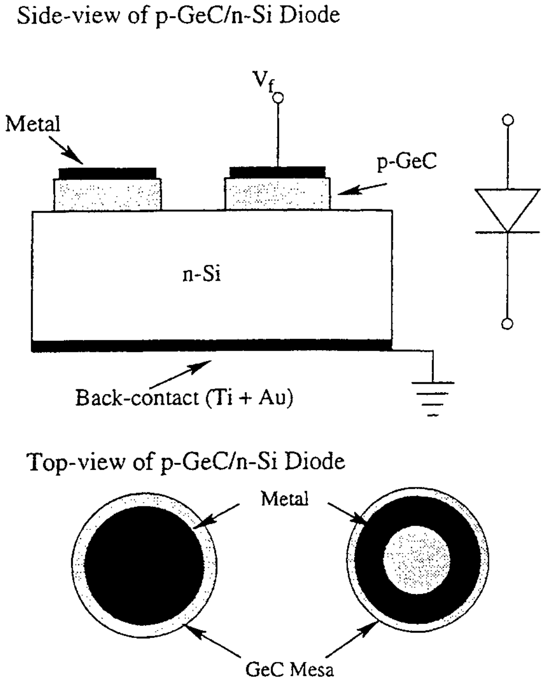
The normal-incidence heterojunction p-n diodes were fabricated from $Ge_{1-x}C_x$ epilayers grown on a (100) n-Si wafer by MBE in an EPI 620 system. Details of the $Ge_{1-x}C_x$ growth are described elsewhere.[8] The carrier concentration of the n-Si substrate is 10^{14} - 10^{15} cm^{-3}. The substrate temperature during growth was kept at 400°C. This temperature produces layer-by-layer growth of GeC to avoid growth front roughness, which can be created at elevated temperatures. At higher temperatures, growth takes place by a Stranski-Krastanov growth mode, resulting in an island morphology. Also, this growth temperature minimizes outdiffusion of the B dopant into the Si substrate. The $Ge_{1-x}C_x$ epilayers were *in situ* doped p-type by a concurrent boron flux, using an effusion cell loaded with pure boron in a pyrolytic graphite crucible. The carbon concentrations were determined by the growth condition, calibrated by C-resonant RBS studies of other higher C concentration samples.[11]

Our first p-n diode was fabricated from a p-$Ge_{0.998}C_{0.002}$ epilayer on a n-Si substrate. The epilayer was a relaxed single crystal with a high density of defects, including misfit and threading dislocations, as expected for epilayers greatly exceeding the critical thickness. This was confirmed by transmission electron microscopy (TEM) in cross-sectional views. The epilayer is continuous and uniform with a flat growth front and the interface between the GeC epilayer and Si substrate is abrupt. The TEM cross-section specimen was prepared by mechanical thinning followed by Ar ion milling at 4 kV. The thickness of the epilayer is 0.6 μm, also measured from a TEM cross-sectional sample. From Hall effect measurements, the electrically active B concentration was about 4×10^{18} cm^{-3} for this epilayer. Additionally, three more $Ge_{1-x}C_x$ epilayers with carbon concentrations of 0.8%, 1.4%,and 2%, were grown, each with a thickness of 0.86 μm and B concentration of 4.4×10^{18} cm^{-3}, as determined from the growth conditions.

A schematic of the device is shown in Fig. 1. Circular GeC mesas were formed by first defining circular regions on the GeC epilayer by photolithography and then etching the uncovered parts of the epilayer down to the Si substrate, using a H_3PO_4:H_2O_2:H_2O etchant solution.[12] The diameter of the mesas were 106 μm for the $Ge_{0.998}C_{0.002}$ p-n diode, and 97 μm for the $Ge_{0.992}C_{0.008}$, $Ge_{0.986}C_{0.014}$, and $Ge_{0.98}C_{0.02}$ p-n diodes. Next, a second photolithography step defined the contact window, using standard liftoff technology to form ohmic contacts on top of the GeC mesas. The contact for the $Ge_{0.998}C_{0.002}$ diode was a solid circle with diameter of 88 μm, while for the other three diodes the contacts were rings with inner diameter of 54 μm and outer diameter of 94 μm. Au (for $Ge_{0.998}C_{0.002}$ diode) and Ti/Au bilayer (for all other diodes) contacts were deposited by electron beam evaporation on the front surface. The backside ohmic contact on the Si substrate was also formed by e-beam evaporation of Ti/Au on the entire backside of the substrate. The samples then underwent heat treatments in a forming gas (15% H_2-N_2) ambient. The $Ge_{0.998}C_{0.002}$ sample was annealed at 350°C for 3 minutes in an annealing furnace, while $Ge_{0.992}C_{0.008}$, $Ge_{0.986}C_{0.014}$, and $Ge_{0.98}C_{0.02}$ samples were annealed at 300°C for 30 seconds in a Heatpulse 610 rapid thermal annealing (RTA) furnace. Adjacent to the diodes were transmission line method (TLM) test structures for measuring the ohmic contact resistance between the top surface metalization and the $Ge_{1-x}C_x$ epilayer. The metal contacts were determined to be ohmic with a contact resistance of only 5.6×10^{-6} $\Omega\cdot cm^2$ for Au[10] and under 2×10^{-6} $\Omega\cdot cm^2$ for Ti/Au.

3. RESULTS AND DISCUSSION

The I-V characteristics of the p-GeC/n-Si diode were measured and are shown in Fig. 2, 3, and Fig. 4. The GeC/Si structures exhibit diode rectification. Ohmic behavior of Au/p-GeC and Ti-Au/p-GeC


Side-view of p-GeC/n-Si Diode

Top-view of p-GeC/n-Si Diode

Figure 1: A schematic of the GeC/Si p-n diode.

was confirmed by contact measurements of Au/p-GeC and Ti-Au/p-GeC, as discussed above. The Ti-Au/n-Si substrate junction is also expected to be ohmic in nature, since the contact area is over the full surface. This results in a very low current density which even if Schottky-like would simulate an ohmic contact through defect-related tunneling on the unpolished substrate backside. In addition, a p-GeC/n-Si diode contacted in series to a n-Si/Ti-Au Schottky diode would behave as two back-to-back diodes. The measured I-V characteristics do not support this supposition. Therefore, diode rectification is believed to be solely influenced by the p-n GeC/Si heterojunction.

The I-V curves, shown in Fig. 2, and 3, exhibit a low turn-on voltage of only 0.15 $\sim$ 0.2 V for all the p-n GeC/Si heterojunctions regardless of carbon composition. At low bias, the measured ideality factor for the $Ge_{0.998}C_{0.002}$ diode is about 1.29. A noticeable feature in the forward I-V characteristics of Fig. 2 is a slight s-shape, or kink, under small forward bias. The serpentine I-V curve suggests that some tunneling current is taking place but which is superimposed upon a much larger forward biased diffusion current. Indeed, TEM analysis reveals that the interface between the $Ge_{0.998}C_{0.002}$ epilayer and Si substrate is abrupt with visible strain contrast and a large quantity of defects and dislocations which may lead to a defect related tunneling current component. The reverse breakdown voltage for the $Ge_{0.998}C_{0.002}$ device, shown in Fig. 2, is in excess of -40 volts. The reverse saturation current for this device is 1.2 nA/μm^2 at -30 volts, and falls to 89 pA/μm^2 at -1 volt.

The three later diodes with improved geometry design, $Ge_{0.992}C_{0.008}$, $Ge_{0.986}C_{0.014}$, and $Ge_{0.98}C_{0.02}$, exhibit improved I-V characteristics. The ideality factor for these three devices is about 1.08, 1.07, and 1.05, respectively. The reverse breakdown voltage of these three devices is in excess of -60 volts, and

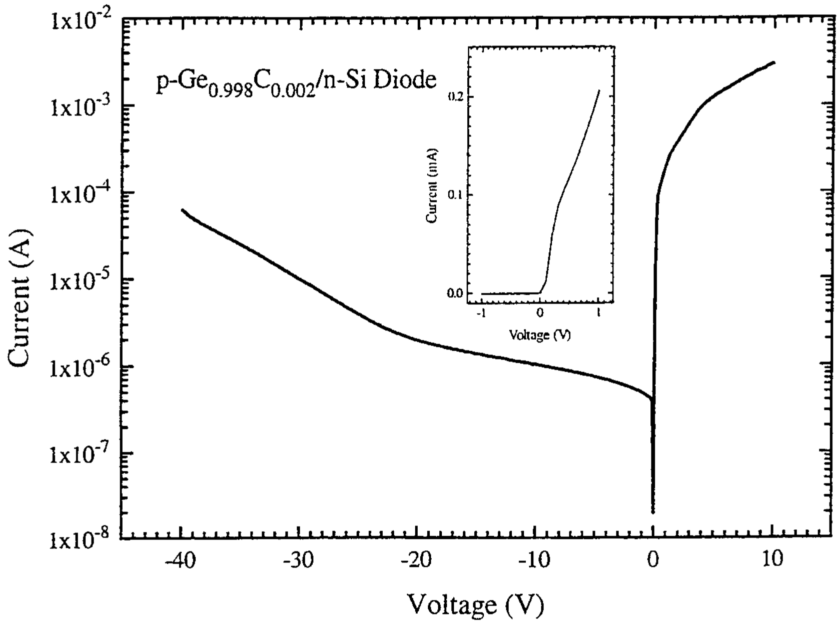


Figure 2: The I-V characteristics of the p-$Ge_{0.998}C_{0.002}$ diode on a n-Si substrate.

some reach -80 volts, irrespective of the C concentration varying from 0.8% to 2%. The measured reverse saturation currents at -20 volts and -1 volts are in the range of 34 - 74 pA/μm^2 and 10 - 14 pA/μm^2, respectively. Again, no obvious dependence on the C concentration was observed within this variation of C concentration. The small leakage current is comparable to a reported leakage current density of 70 pA/μm^2 for a SiGeC *p-i-n* photodetector.[13] The higher breakdown voltage and lower reverse saturation current of the three later diodes could also be due to an improved quality of the three later GeC epilayers which were grown at a slower growth rate than the first GeC layer.

Photoresponsivity was observed for all the GeC/Si heterojunction diodes using laser excitation at a wavelength of 1.3 μm, which is below the Si bandgap (λ = 1.1 μm). For the $Ge_{0.998}C_{0.002}$ diode, the laser spot was focused between the periphery of the mesa and the upper metal contact. A photocurrent of 0.45 μA was measured under an incident light power of 31 μW for this diode, as shown in Fig. 5, which corresponds to a responsivity of 1.45×10^{-2} A/W and an external quantum efficiency (EQE) of 1.4%. The ring-shaped upper metal contacts for the other three annular devices facilitated easier focusing of the laser beam within the inner circular window of the contact. Larger photocurrents were measured in these three samples and are also shown in Fig. 5 and Fig. 6. The photocurrents of the $Ge_{0.992}C_{0.008}$, $Ge_{0.986}C_{0.014}$ and $Ge_{0.98}C_{0.02}$ diodes under applied reverse bias of 10 volts were 4.2, 3.7 and 2.5 μA respectively, corresponding to external quantum efficiencies (EQE) of 2.3%, 2%, and 1.2%, respectively, when the total incident light power of 177 μW was used in the calculation (Fig. 7). If the fact that a fraction of the light falls outside of the active area is considered, the EQE's could be greater than 2.8, 2.5, and 1.5%, respectively, assuming a 2D Gaussian distribution for the light intensity. Although the difference between the quantum efficiencies of the $Ge_{0.992}C_{0.008}$ and $Ge_{0.986}C_{0.014}$ diodes is not obvious, considering experimental errors, a large drop in quantum efficiency was observed for the $Ge_{0.98}C_{0.02}$ diode, which has the largest carbon concentration in

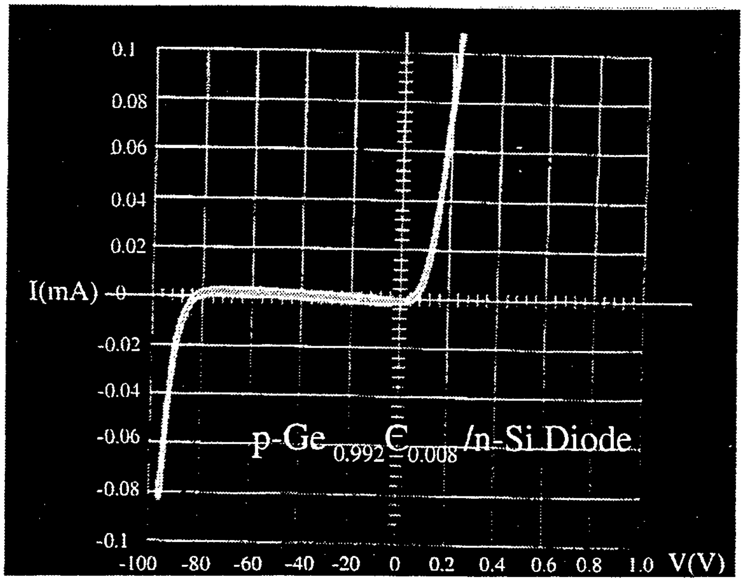


Figure 3: The I-V characteristics of the p-$Ge_{0.992}C_{0.008}$/n-Si diode.

the group. It is possible that there is a non-linear dependence of photoresponsivity on carbon concentration and this dependence could saturate when the carbon fraction is below some certain level. The results of 1.2 - 2.3 % quantum efficiency are significant not only because they are greater than the best result reported in SiGeC *p-i-n* photodetectors, but also because the diodes did not have an anti-reflection (AR) coating and the light collection region is confined to a narrow region at the p^+-GeC/n^--Si junction. This leads to a much shorter path length (a few tens of angstroms) within the p^+-GeC depletion region than previous normal incidence SiGeC *p-i-n* photodetectors which reported EQE ~1% at 1.3 μm, using an 800 $\AA$ thick active region.[13]

4. CONCLUSION

In conclusion, p-n diodes made from a heterojunction of epitaxial p-type $Ge_{1-x}C_x$ with carbon percentages of 0.2, 0.8, 1.4, and 2% on an n-type Si substrate were demonstrated. The I-V curve of the p-GeC/n-Si diodes indicates low reverse saturation current (10 pA/μm^2 at -1 volt) and high reverse breakdown voltage in excess of -40 and some up to -80 volts. Despite the large number of dislocations and defects at the heterojunction, photoresponses with measured EQE of 1.2 - 2.3 % from the $Ge_{1-x}C_x$ epilayer were observed by 1.3 μm laser excitation . It is possible that there is some non-linear dependence of EQE on carbon concentration.

5. ACKNOWLEDGMENTS.

The authors wish to thank F. Chen for Hall measurements. This work was supported by DARPA (Sponsored Research Agreement with Texas Instruments, Inc. under Grant No. SRA-3312665) and the National Science Foundation (ECS-9624160).

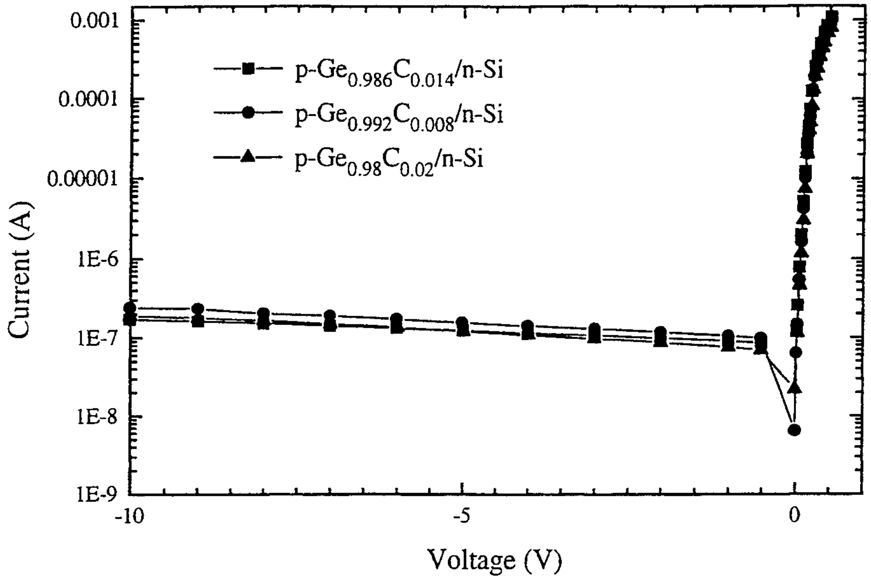


Figure 4: The measured I-V characteristics of p-Ge$_{1-x}$C$_x$/n-Si diodes with 0.8, 1.4, and 2.0% carbon.

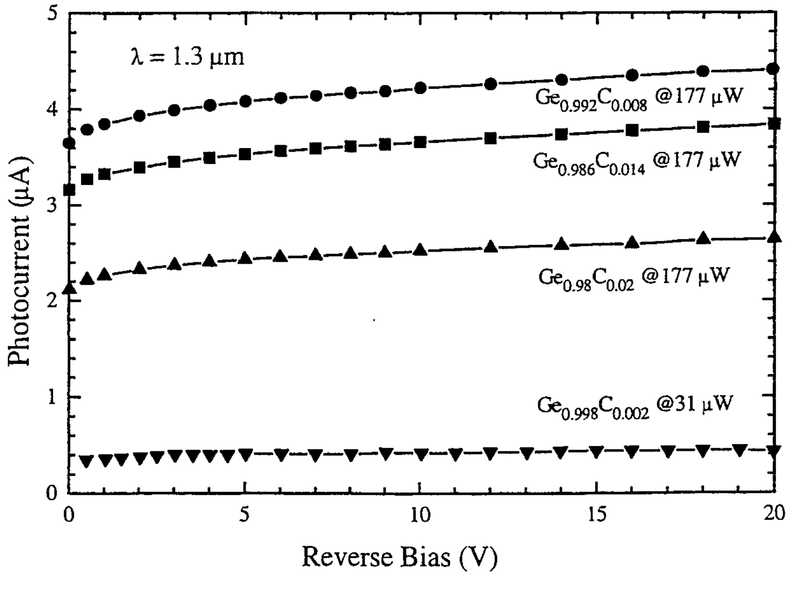


Figure 5: The measured photoresponsivity for the p-Ge$_{1-x}$C$_x$/n-Si diodes with 0.2, 0.8, 1.4, and 2% carbon.

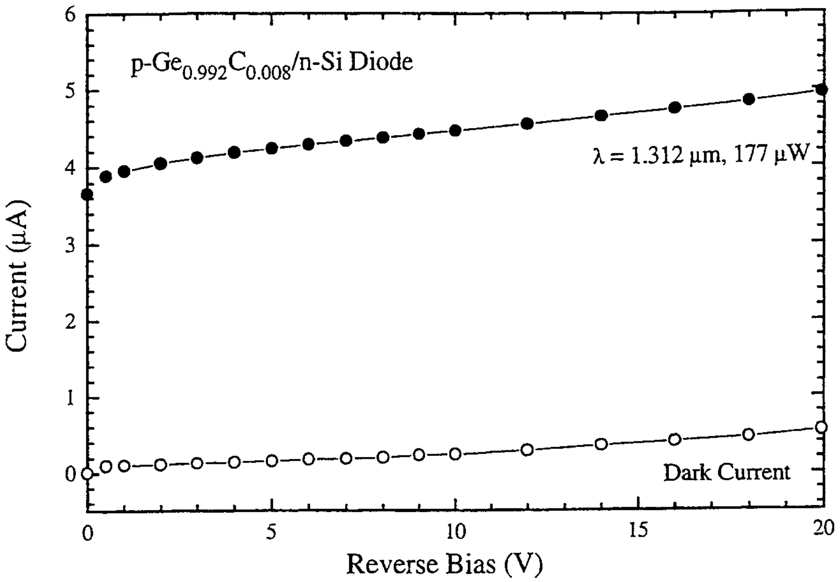


Figure 6: The measured dark current and photocurrent under 177 μW of incident 1.3 μm laser illumination for the p-$Ge_{0.992}C_{0.008}$/n-Si photodiode.

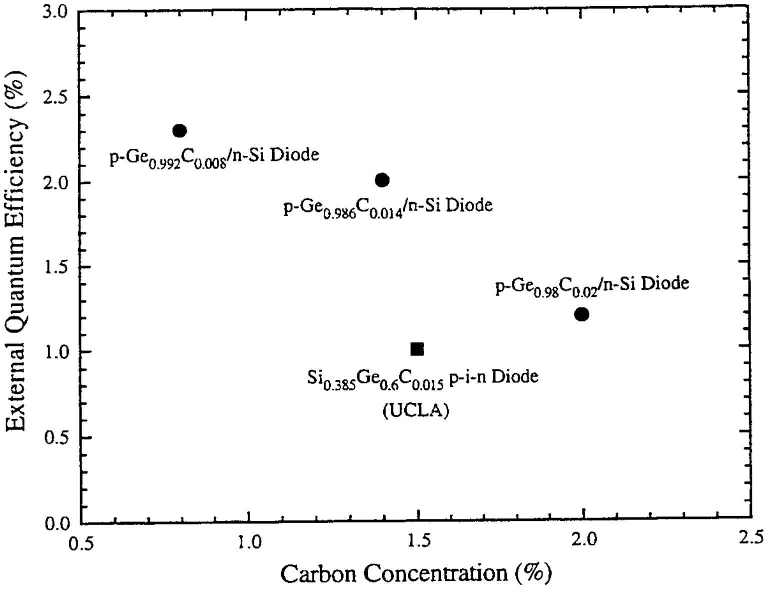


Figure 7: The measured external quantum effeciency for the p-$Ge_{1-x}C_x$/n-Si photodiodes and compared to a SiGeC p-i-n diode [*Appl. Phys. Lett.*, vol. 69, 2330 (1996)].

6. REFERENCES

1. R. A. Soref, "Silicon-based optoelectronics," *Proceedings of the IEEE*, vol. 81, pp. 1687-1706 (1993).

2. R. People, "Physics and applications of Ge_xSi_{1-x}/Si strained layer heterostructures," *IEEE J. of Quantum Electron.*, vol. QE-22, pp. 1696-1710 (1986).

3. K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, "Growth and strain compensation effects in the ternary $Si_{1-x-y}Ge_xC_y$ alloy system," *Appl. Phys. Lett.*, vol. 60, pp. 3033-3035 (1992).

4. B. Dietrich, H. J. Osten, H. Rücker, M. Methfessel, and P. Zaumseil, "Lattice distortion in a strain-compensated $Si_{1-x-y}Ge_xC_y$ layer on silicon," *Phys. Rev. B*, vol. 49, pp. 17185-17190 (1994).

5. F. J. Guarin, S. S. Iyer, A. R. Powell and B. A. Ek, "Growth and strain symmetrization of Si/Ge/C/Sn quaternary alloys by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 68, pp. 3608-3610 (1996).

6. R. I. Scace and G. A. Slack, "Solubility of carbon in silicon and germanium," *J. Chem. Phys.*, vol. 30, pp. 1551-1555 (1959).

7. O. F. Sankey, A. A. Demkov, W. T. Petuskey, and P. F. McMillan, "Energetics and electronic structure of the hypothetical cubic zincblende form of GeC," *Modelling Simul. Mater. Eng.*, vol. 1, pp. 741-754 (1993).

8. J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, "Growth of germanium-carbon alloys on silicon substrates by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 67, pp. 1865-1867 (1995).

9. M. Todd, J. Kouvetakis, and D. Smith, "Synthesis and characterization of heteroepitaxial diamond-structured $Ge_{1-x}C_x$ (x=1.5-5.0%) alloys using chemical vapor deposition," *Appl. Phys. Lett.*, vol. 68, pp. 2407-2409 (1996).

10. X. Shao, S. L. Rommel, B. A. Orner, P. R. Berger, and J. Kolodzey, "Low resistance ohmic contacts to p-$Ge_{1-x}C_x$ on Si," *IEEE Electron. Dev. Lett.*, vol. 18, pp. 7-9 (1997).

11. D. A. Hits, B. A. Orner, J. O. Olowolafe, J. Kolodzey, T. Laursen and J. W. Mayer, "Location of C atoms in the structure of Ge-rich $Si_{1-x-y}Ge_xC_y$ on Si," *submitted to Appl. Phys. Lett.*

12. A-S. T. Khan, *Photoluminescence and process issues of SiGeCSn alloys*, Ph.D. Thesis, University of Delaware, (1996).

13. F. Y. Huang and K. L. Wang, "Normal-incidence epitaxial SiGeC photodetector near 1.3 μm wavelength grown on Si substrate," *Appl. Phys. Lett.*, vol. 69, pp. 2330-2332 (1996).

Optical constants of B and P doped $Ge_{1-y}C_y$ alloys on Si substrates

B. A. Orner, F. Chen, D. Hits, M. W. Dashiell, and J. Kolodzey

Department of Electrical Engineering
University of Delaware, Newark, DE 19716

ABSTRACT

$Ge_{1-y}C_y$ alloys are meta-stable and challenging to produce due to the large disparity between the atomic sizes of Ge and C. However, this same disparity results in an alloy system that potentially spans a wide range of bandgaps, refraction indices, and lattice constants. As such, it has potential as a Si lattice matched material for use in Si based waveguides, detectors, modulators, and other devices. The performance of these devices, however, depends on the refractive indices, which are not well known in these alloys. We present the results of comprehensive measurements of refractive index, energy bandgap, and free carrier absorption versus doping level. *In situ* B and P doped 550 nm $Ge_{1-y}C_y$ alloy films were grown on Si (100) substrates by molecular beam epitaxy. The infrared optical transmission spectra were measured at room temperature. The indices of refraction were obtained from the amplitude of the interference fringes at sub-bandgap photon energies. Hall effect measurements were employed to measure the carrier concentrations. The refractive index of our $Ge_{1-y}C_y$ alloys was nominally 4.01, and decreased with increasing B and P concentration.

Keywords: absorption, refraction, epitaxy, group IV alloys

1 INTRODUCTION

Binary and ternary alloys of Si, Ge, and C are being investigated for use in heterostructure optical, electrical, and electro-optical devices.[1,2] These alloys may be epitaxially grown on Si and Ge substrates, and some compositions may be lattice matched to Si. Alloys containing C are meta-stable but may be produced by non-equilibrium epitaxial growth techniques such as molecular beam epitaxy (MBE),[3,4] chemical vapor deposition (CVD),[5,6] and solid phase epitaxy (SPE).[7]

Currently, little is known about the optical properties of group IV alloys containing C. The alloys are expected to be indirect band gap materials at all compositions,[8] and the band gap energy is likely to vary non-linearly with composition,[9] though further study is necessary.

Many heterostructure optical and electro-optical devices based on these materials have been proposed, including photodiodes, modulators, waveguides, self electro-optic effect devices, Bragg reflectors, and emitters.[1] The ability to dope the alloys to control the electrical and optical properties is crucial for many devices. Yet doping is problematic in many new materials, and few papers have addressed doping in group IV alloys containing C.

2 EPITAXIAL GROWTH

In this paper we demonstrate through refractive index changes and Hall effect measurements the p-type and n-type doping of $Ge_{1-y}C_y$ alloys epitaxially grown by MBE on Si substrates. The substrates were (100) oriented, single side polished substrates with resistivities of 1 to 10 Ω-cm. For p-type B doped layers, the substrates were n-type, and for n-type P doped layers, the substrates were p-type.

The films were ***in situ*** doped with B and P from solid source effusion cells. For p-type doping, B was evaporated in a pyrolytic graphite (PG) crucible at 1450 to 1650 °C. Phosphorus doping for n-type layers was accomplished by evaporation of GaP in a pyrolytic boron nitride (PBN) crucible at 600 to 750 °C. Phosphorus is preferentially evaporated at these temperatures, and a PBN gettering aperture in the molecular beam above the crucible further reduces Ga.[10] Zone refined solid Ge in a PBN crucible was used as the Ge source, and a PG filament served as the C source. The films were approximately 0.6 microns thick as measured by a mechanical stylus. The C content was estimated from the growth conditions to be 0.12 %. Reciprocal lattice rods were observed by reflection high energy electron diffraction (RHEED) after growth, suggesting crystalline films. The alloys were well above the critical thickness for Ge rich alloys on Si, and were thus assumed to be relaxed.

3 ANALYSIS

3.1 Carrier concentration

During Hall effect measurements, the p-n junction formed by the substrate and the thin film was reverse biased to electrically isolate the epitaxial film from the substrate. Room temperature Hall effect measurements were then performed to provide the carrier concentration in the alloy films.[11] The carrier concentration correlated well with the dopant cell temperatures. For B doping, the hole concentrations were 6.7×10^{17} (sample number SGC-81), 1.7×10^{18} (SGC-80), and 1.0×10^{19} (SGC-82). For P doping, the electron concentrations were 1.4×10^{18} (SGC-157), 2.3×10^{18} (SGC-158), and 1.2×10^{19} (SGC-159).

3.2 Optical Transmission

The transmission through the samples in the near infrared region was measured at room temperature by a Fourier transform infrared spectrometer (FTIR).[12] The transmission spectra had three principal components: substrate absorption, alloy film absorption, and interference fringes resulting from multiple reflections at the air/film and film/substrate interface. In the region of interest for the present study (energies at or below the alloy band gap energy), the substrate absorption was minimal and removable. The interference fringes were analyzed to obtain the refractive index of the alloy, and the alloy film absorption was examined to determine the alloy band gap energy.

Below the substrate band gap energy (1100 meV) the absorption was small and constant. The absorption in this region was removed from the spectra by ratioing the sample spectra to a substrate reference. Above the substrate band gap energy substrate absorption prevented transmission measurements.

The transmission spectra was given by:

$$T = C\frac{(t_1t_2)^2e^{-\alpha d}}{1+(r_1r_2)^2e^{-2\alpha d}+2r_1r_2e^{-\alpha d}\cos(2\phi-\pi)} \tag{1}$$

where r_1, t_1, r_2, and t_2 are the Fresnel reflection and transmission coefficients of the alloy/air and alloy/substrate interfaces, and d is the alloy layer thickness. C, a factor to correct for experimental optical losses, was near unity. In this case the Fresnel coefficients are:

$$r_1 = \frac{1-n}{1+n}, \tag{2}$$

$$r_2 = \frac{n-n_{sb}}{n+n_{sb}}, \tag{3}$$

$$t_1 = \frac{2}{1+n}, \tag{4}$$

$$t_2 = \frac{2n}{1+n_{sb}}, \tag{5}$$

and

$$\phi = 2\pi nd(1/\lambda) + \psi. \tag{6}$$

In the above equations, n and n_{sb} are the refractive indices of the $\mathrm{Ge}_{1-y}\mathrm{C}_y$ layer and the substrate respectively, d is the alloy layer thickness, and λ is the vacuum wavelength of the incident photons. A phase correction factor ψ was also included to obtain an improved curve fit.

3.3 Refractive index

Below the alloy band gap, the alloy absorption was assumed to be negligible, and Eq. 1 reduced to:

$$T = \frac{(t_1 t_2)^2}{1 + (r_1 r_2)^2 + 2r_1 r_2 \cos(2\phi - \pi)}. \tag{7}$$

The spectra thus consisted of interference fringes, the amplitude of which was determined by the refractive indices of the substrate and the alloy. This provided a sensitive measure of the index of refraction of the alloy films. The index of refraction was assumed constant with respect to energy below the band gap of each material, and the refractive index of the substrate was assumed to be 3.45. Curve fitting was employed to fit Eq. 7 to the spectra in the region between 500 and 710 meV (Fig. 1). Adjustable fitting parameters included n, d, C, and ψ.

The relation between the carrier concentration and the refractive index obtained from the curve fit is shown in Fig. 2. The refractive index decreased with decreasing carrier concentration. The reduction in the refractive index is related to the increased carrier concentration by:[13]

$$n = n_0 - \frac{Nq^2}{2n_0 \epsilon_0 m^* \omega^2}, \tag{8}$$

where n_0 is the refractive index of intrinsic material, N is the extrinsic carrier concentration, m^* is the conductivity effective mass, ω is the incident photon frequency, ϵ_0 is the electric permittivity, and q is the magnitude of the electron charge.

We fixed ω at the value for 650 meV photons (9.86×10^{14} rad/s), and fit Eq. 8 to our refractive index data. The respective values for holes in Ge are $n_0 = 4.0$ and $m^* = 0.23m_0$.[14] Our results for the intrinsic refractive index are similar to Ge, as expected. Our results suggest that in this composition range $\mathrm{Ge}_{1-y}\mathrm{C}_y$ alloys have a lower effective hole mass than Ge, which is consistent with the increased mobility we have observed in these alloys.[11] The fit for the P doped samples indicated $n_0 = 4.04$ but had too much residual error to determine m^*.

3.4 Fundamental absorption

We obtained α from T by solving Eq. 1. This removed the interference fringes from the spectra, resulting in absorption coefficient spectra (Figs. 3-4). The fundamental absorption edge was observed to shift with doping

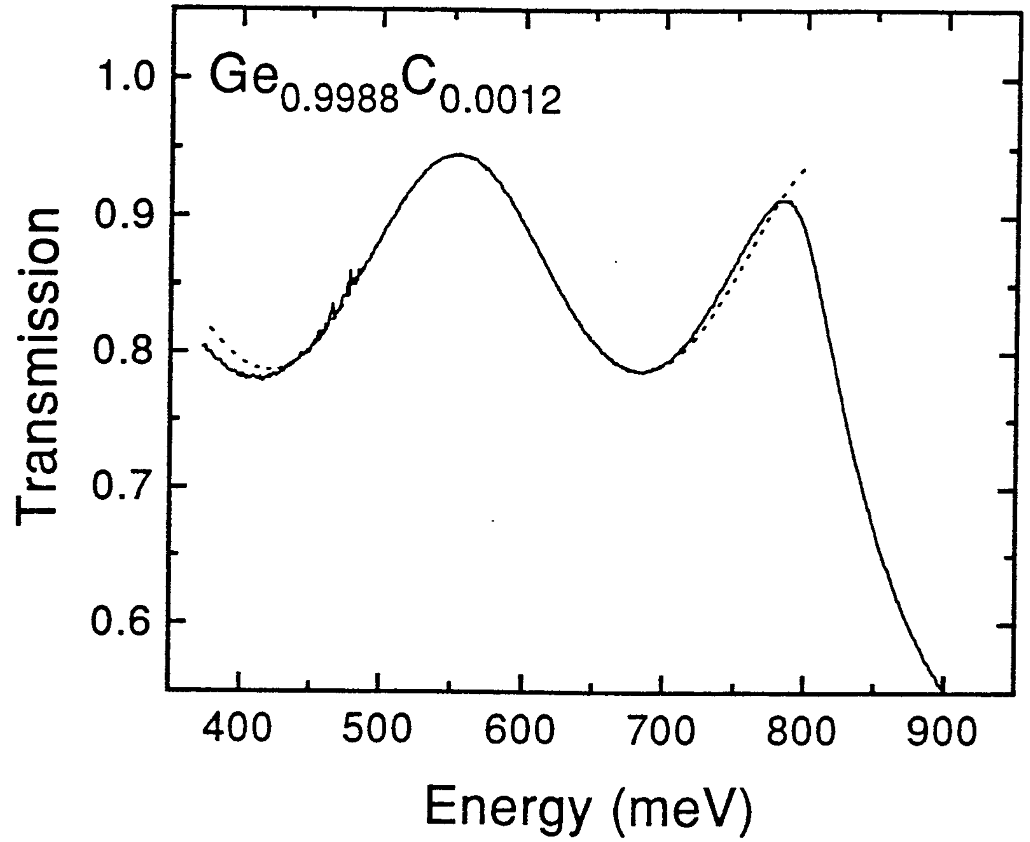


Figure 1: Optical transmission (solid line) of a 550 nm $Ge_{0.9988}C_{0.0012}$ B doped thin film on a Si (100) substrate. The interference fringes resulted from multiple reflections within the alloy film. We determined the index of refraction and layer thickness by curve fitting (dashed line) to these fringes.

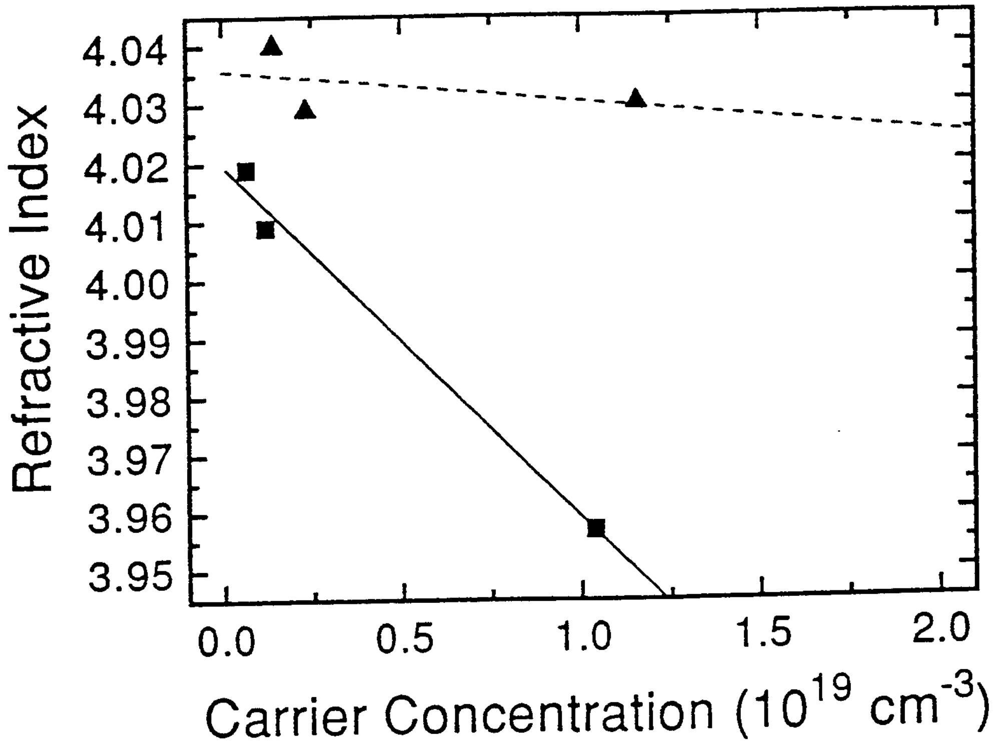


Figure 2: Refractive index versus experimental hole concentration (squares) and experimental electron concentration (triangles) for $Ge_{0.9988}C_{0.0012}$ alloys. The refractive index was determined from curve fitting to thin film interference fringes, and the carrier concentration was determined from Hall effect measurements. Theoretical fits to the data are illustrated for holes (solid line) and electrons (dashed line).

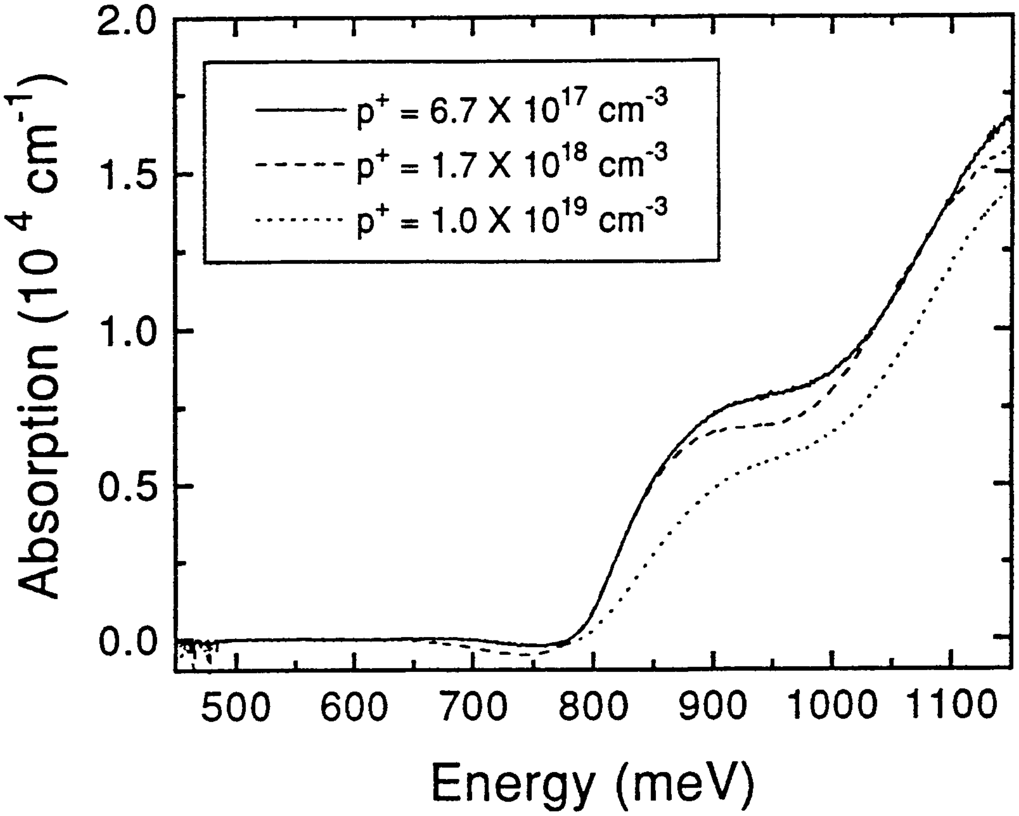


Figure 3: Measured fundamental optical absorption in B doped $Ge_{0.9988}C_{0.0012}$. A blue-shift is observed with increasing carrier concentration. Oscillations are artifacts due to imperfect cancelation of interference fringes.

level. Increased doping often causes a decrease in the effective band gap energy (bandgap narrowing) due to the addition of shallow donor or acceptor impurity states at the edges of the band gap.[14] Doping can also cause an increase in the effective optical band gap (Burstein shift) if dopant levels are sufficient to move the Fermi level out of the band gap.[15] In this case, states near the edge of the conduction band of an n-type material are occupied by electrons, and similarly states near the edge of the valence band of a p-type material are occupied by holes.

We observed a blue-shift in the fundamental absorption edge of the B doped samples (Fig. 3) as the carrier concentration increased. For the P doped samples (Fig. 4), we observed a red-shift in the fundamental absorption edge as the carrier concentration increased. Although we are not yet sure of the mechanism for the different directions of shift with doping, we speculate that the density-of-states effective mass of the holes may be lower than that of the electrons. That may lead to Burstein shifts dominating for holes and to bandgap narrowing dominating for electrons.

The Macfarlane-Roberts expression[12,16] for fundamental absorption in an indirect semiconductor may be fit to

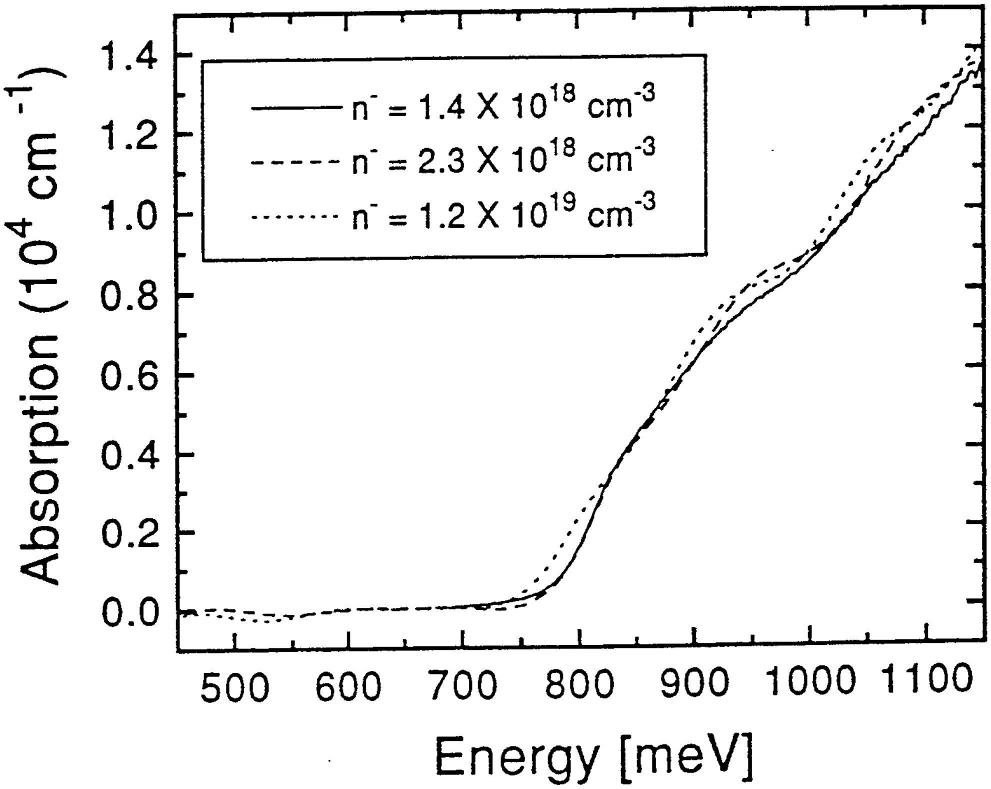


Figure 4: Measured fundamental optical absorption in P doped $Ge_{0.9988}C_{0.0012}$. A red-shift is observed with increasing carrier concentration.

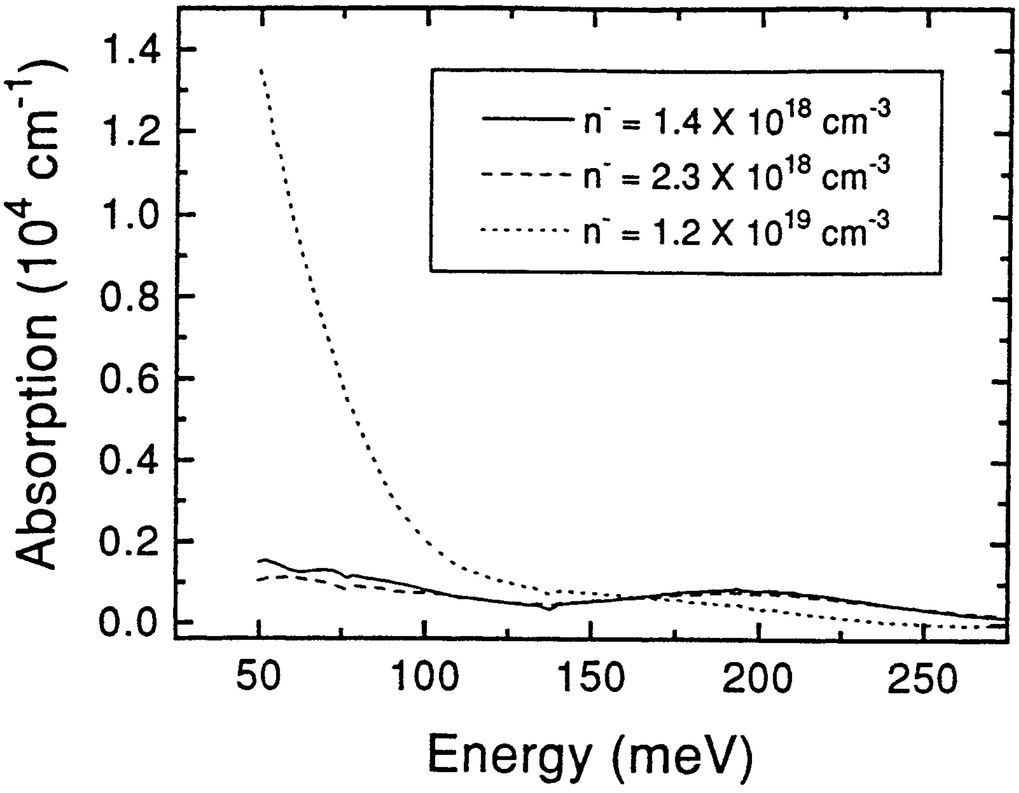


Figure 5: Mid-infrared absorption of P doped $Ge_{0.9988}C_{0.0012}$ alloy films. The higher absorption below 100 meV in the sample doped to 1.2×10^{19} was attributed to free carrier absorption.

the absorption curves of Figs. 3-4. This procedure was complicated by imperfect interference fringe removal and possible direct electron-hole transitions at the Γ point at energies near the minimum indirect band gap energy. Therefore, experimental uncertainties were on the order of the observed shift in the band gap energy with doping. The band gap energies obtained from curve fitting varied from 647 to 686 meV, which agrees with expectations[3] for a Ge rich $Ge_{1-y}C_y$ alloy.

3.5 Free carrier absorption

The absorption at lower photon energies was measured to study the free carrier absorption in this material. The results for the P doped samples are shown in Fig. 5; the results for B doped samples were similar. The sample with an electron concentration of 1.16×10^{19} cm^{-3} had significantly higher free carrier absorption. The free carrier absorption was most prevalent at photon energies below 100 meV.

4 CONCLUSION

In conclusion, we have demonstrated tuning of the optical properties of $Ge_{1-y}C_y$ through B and P doping. Incorporation of activated donor and acceptor impurities was demonstrated by Hall effect studies, a decreased refractive index, and changes to the fundamental absorption edge. We have demonstrated successful donor and acceptor doping of this material. The ability to employ carrier concentrations to tune the refractive index is useful for certain devices such as optical waveguides and modulators.

5 ACKNOWLEDGMENTS

This work was supported by ONR grant N00014-93-1-0393 and DARPA contract F49620-96-C-0006.

6 REFERENCES

[1] R. A. Soref. Silicon-based optoelectronics. *Proc. IEEE*, 81(12):1687–1706, 1993.

[2] A. R. Powell, K. Eberl, F. E. LeGoues, B. A. Ek, and S. S. Iyer. Stability of strained $Si_{1-y}C_y$ random alloy layers. *J. Vac. Sci. Technol.*, B11(3):1064–8, 1993.

[3] J. Kolodzey, P. A. O'Neil, S. Zhang, B. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite, and S. Ismat Shah. Growth of Germanium-Carbon alloys on Silicon substrates by molecular beam epitaxy. *Appl. Phys. Lett.*, 67(13):1865–1867, 1995.

[4] H. J. Osten and J. Klatt. *In situ* monitoring of strain relaxation during antimony-mediated growth of Ge and $Ge_{1-y}C_y$ layers on Si(001) using reflection high energy electron diffraction. *Appl. Phys. Lett.*, 65:630, 1994.

[5] P. Boucaud, C. Francis, F. H. Julien, J. M. Lourtoiz, D. Bouchier, S. Bodnar, B. Lambert, and J.L. Regolini. Band-edge and deep level photoluminescence of pseudomorphic $Si_{1-x-y}Ge_xC_y$ alloys. *Appl. Phys. Lett.*, 64(7):875–877, 1994.

[6] A. St. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewalt. Defect-free band-edge photoluminescence and band gap measurement of pseudomorphic $Si_{1-x-y}Ge_xC_y$ alloy layers on Si (100). *Appl. Phys. Lett.*, 67(26):3915–3917, 1995.

[7] J. W. Strane, H. J. Stein, S. R. Lee, B. L. Doyle, S. T. Picraux, and J. W. Mayer. Metastable SiGeC formation by solid phase epitaxy. *Appl. Phys. Lett.*, 63(20):2786–2788, 1993.

[8] R. A. Soref. Optical band gap of the ternary semiconductor $Si_{1-x-y}Ge_xC_y$. *J. Appl. Phys.*, 70(4):2470–2472, 1991.

[9] A. A. Demkov and O. F. Sankey. Theoretical investigation of random Si-C alloys. *Phys. Rev.*, B48:2207, 1993.

[10] G. Lippert, H. J. Osten, and D. Kruger. Phosphorous doping in molecular-beam epitaxial grown silicon and silicon-germanium using a GaP decomposition source. *J. Crystal Growth*, 157(1-4):304–307, 1995.

[11] F. Chen, K. Roe, D. Hits, B. A. Orner, D. S. Holmes, R. G. Wilson, and J. Kolodzey. Electrical properties of $Si_{1-x-y}Ge_xC_y$ and $Ge_{1-y}C_y$ Alloys. *J. of Elec. Materials*, In Press, 1997.

[12] B. A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J. W. Mayer, and J. Spear. Bandgap of Ge rich $Si_{1-x-y}Ge_xC_y$ alloys. *Appl. Phys. Lett.*, 69(17):2557–2559, 1996.

[13] Robert G. Hunsperger. *Integrated Optics: Theory and Technology.* Springer-Verlag, 4th edition, 1995.

[14] M. S. Tyagi. *Introduction to Semiconductor Materials and Devices.* John Wiley and Sons, 1991.

[15] K. Seeger. *Semiconductor Physics.* Springer-Verlag, 1973.

[16] G. G. Macfarlane, T. P. McLean, J. E. Quarrington, and V. Roberts. Fine structure in the absorption-edge spectrum of Ge. *Phys. Rev.*, 108(6):1377–1383, 1957.

IEEE ELECTRON DEVICE LETTERS, VOL. 18, NO. 9, SEPTEMBER 1997

A p-$Ge_{1-x}C_x$/n-Si Heterojunction Diode Grown by Molecular Beam Epitaxy

Xiaoping Shao, S. L. Rommel, *Student Member, IEEE*, B. A. Orner, *Student Member, IEEE*,
J. Kolodzey, *Senior Member, IEEE*, and Paul R. Berger, *Senior Member, IEEE*

Abstract—**We report on the fabrication and characterization of the first p-n diode made from a heterojunction of epitaxial p-type $Ge_{0.998}C_{0.002}$ on an n-type Si substrate. Epitaxial $Ge_{0.998}C_{0.002}$ was grown on a (100) Si substrate by solid source molecular beam epitaxy. The p-GeC/n-Si junction exhibits diode rectification. The I-V characteristics of the p-GeC/n-Si diode indicate a reasonable reverse saturation current of 89 pA/μm^2 at -1 V and a high reverse breakdown voltage in excess of -40 V. Photoresponse from the $Ge_{0.998}C_{0.002}$ p-n diode was observed from 1.3-μm laser excitation resulting in an external quantum efficiency of 1.4%.**

I. INTRODUCTION

GROUP IV semiconductor alloys have attracted intense studies over the last decade for their potential applications in electronic as well as optoelectronic integrated circuits on Si [1]–[2]. Due to the 4.2% lattice mismatch between Si and Ge, most investigations have involved SiGe heterostructures with limited Ge concentration ($\leq$30%) to enhance device performance [3]. The formation of the ternary alloy $Si_{1-x-y}Ge_xC_y$, by adding C to the SiGe system, makes it possible to reduce the strain within the SiGeC system about the lattice matching condition for Si substrates while allowing adjustable bandgaps [4]–[6] by altering the Ge:C ratio. The binary alloy $Ge_{1-x}C_x$ also provides the same flexibility from pure Ge under compressive strain (4.2% misfit) to diamond under tensile strain (-34.3% misfit) and has been studied by a number of groups [7], [8]. Osten *et al.* [7] showed that inclusion of C into Ge delays the onset of strain relaxation, and does not behave identically to Ge with an artificially reduced strain.

Group IV alloy materials research is beginning to spawn device applications, such as strain compensated SiGeC base layers in heterojunction bipolar transistors [9]. Our own investigations have begun to explore materials issues [10] and processing issues [11] of the GeC binary alloy, and have now expanded to investigate novel devices. In this letter, we report on the fabrication and characterization of the first p-n diode made from a heterojunction of molecular beam epitaxial (MBE) p-type $Ge_{1-x}C_x$ epilayer with 0.2% carbon on an n-type Si substrate which demonstrates rectification and photoresponsivity.

One issue which hinders the investigation of Group IV alloys, however, is the limited solubility of C in Ge and Si. Scace and Slack [12] reported the solubility of C in Ge as 10^8 atoms/cm^3 at the melting point of Ge. Also, a theoretical investigation by Sankey et al. [13] approximation indicated that GeC is thermodynamically unstable in solid state form, decomposing into its segregated components, under zero pressure. But, far from thermodynamic-equilibrium growth techniques have been successful in synthesizing $Ge_{1-x}C_x$ alloys with small percentages of carbon, like MBE (under 3% C) [8] or chemical vapor deposition (CVD) with highly reactive precursors (under 5% C) [14].

II. EXPERIMENTS

The $Ge_{1-x}C_x$ epilayer was grown on a (100) n-Si wafer with a carrier concentration of 10^{15} cm^{-3} by MBE in an EPI 620 system without a buffer layer. Details of the $Ge_{1-x}C_x$ growth are described elsewhere [8]. The substrate temperature during growth was kept at 400 °C to minimize outdiffusion of the B dopant into the Si substrate. The measured thickness of the GeC epilayer was 0.6 μm. The C concentration was 0.2%, as determined by the growth condition, calibrated from samples with higher C concentrations.

The $Ge_{0.998}C_{0.002}$ epilayer was doped p-type by a concurrent boron flux, using an effusion cell loaded with pure boron in a pyrolytic graphite crucible during the MBE growth. From Hall effect measurements, the electrically active B concentration was about 4×10^{18} cm^{-3}. The $Ge_{1-x}C_x$ epilayer was also examined in a Philips 400T transmission electron microscope (TEM) in cross sectional views. The TEM specimen was prepared by mechanical thinning followed by Ar ion milling at a voltage of 4 kV.

Heterojunction p-n diodes were fabricated and a schematic is shown in the insert of Fig. 1. Circular GeC mesas were formed by photolithography and then etching down to the Si substrate, using a $H_3PO_4 : H_2O_2 : H_2O$ etchant solution [15]. The diameter of the p-n diode mesas was 106 μm. Standard liftoff technology was used to form e-beam deposited Au ohmic contacts on the GeC mesas. The n-Si ohmic contact was formed by e-beam evaporation of Ti/Au on the entire backside of the substrate. The samples were annealed at 350 °C in a forming gas (15% H_2-N_2) ambient for 3 min. The upper Au contacts were determined to be ohmic [11] using transmission line method (TLM) test structures with a measured contact resistance of only 5.6×10^{-6} Ω·cm^2.

Manuscript received April 20, 1997; revised May 23, 1997. This work was supported by the National Science Foundation (ECS-9624160), AFOSIT (Grant F49620-95-1-0135), and by DARPA (Sponsored Research Agreement with Texas Instruments, Inc. under Grant SRA-3312665).

The authors are with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA.

Publisher Item Identifier S 0741-3106(97)06673-1.

0741–3106/97$10.00 © 1997 IEEE

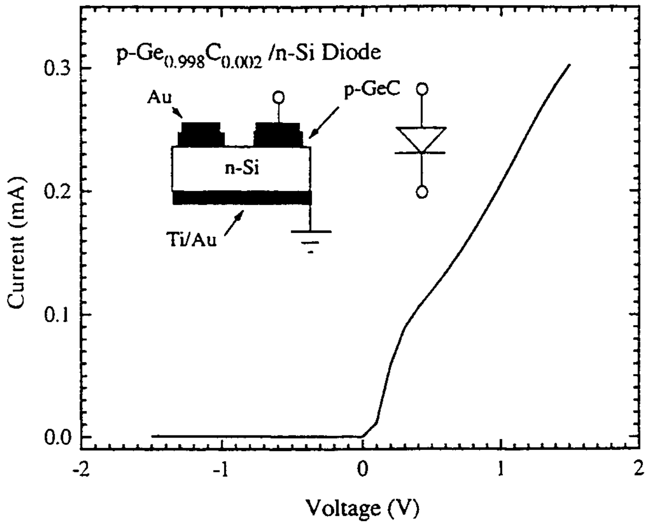


Fig. 1. The measured $I-V$ characteristics of the p-Ge$_{0.998}$C$_{0.002}$/n-Si diode at reduced bias. The insert depicts the fabricated diode structure used in this study.

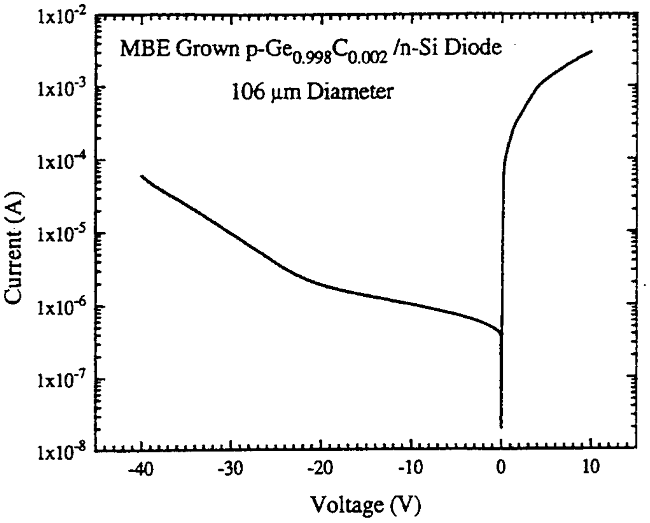


Fig. 2. The measured log $I-V$ characteristics of the p-Ge$_{0.998}$C$_{0.002}$/n-Si diode.

III. RESULTS AND DISCUSSION

The GeC epilayer on Si (~4% misfit) exceeds the critical thickness, which is only a few monolayers. TEM analysis reveals that the epitaxial Ge$_{0.998}$C$_{0.002}$ used in this study is a single crystal. The epilayer is continuous and uniform with a flat surface, and a high density of dislocations threading to the free surface. The interface between the GeC epilayer and Si substrate is abrupt with visible strain contrast and a large quantity of defects and dislocations. Similar, thick dislocated Ge$_{1-x}$C$_x$ epilayers on Si have demonstrated bandedge luminescence [10].

The $I-V$ characteristics of the p-GeC/n-Si diode were measured and are shown in Figs. 1 and 2. The GeC/Si structures exhibit diode rectification. There are three possibilities for the rectified $I-V$ characteristics: 1) a Schottky Au/p-GeC contact on the top surface, 2) the p-GeC/n-Si heterojunction itself, and/or 3) the n-Si/Ti-Au contact on the backside of the substrate. But, measurements confirm ohmic behavior of the Au/p-GeC junction [11], as discussed above, and the n-Si/Ti-Au backside contact is also expected to be ohmic, since the contact area is over the full surface. This results in a very low current density which even if Schottky-like would simulate an ohmic contact through defect-related tunneling on the unpolished substrate backside. In addition, a p-GeC/n-Si contacted in series to a n-Si/Ti-Au Schottky would behave as two back-to-back diodes. The measured $I-V$ characteristics do not support this supposition. Therefore, diode rectification is believed to be solely influenced by the p-GeC/n-Si heterojunction.

The $I-V$ curve, shown in Fig. 1, exhibits a low turn-on voltage of only 0.15–0.2 V. At low bias, the measured ideality factor is about 1.29. As the bias voltage is increased, see Fig. 2, the $I-V$ curve bends over to a smaller slope, which is probably caused by the high series resistance associated with the low-doped n-Si substrate. There is also a slight s-shape, or kink, under small forward bias, see Fig. 1. The serpentine $I-V$ curve suggests that some tunneling current is taking place but which is superimposed upon a much larger forward biased diffusion current. The reverse saturation current, shown in Fig. 2, is reasonable, and exhibits a reverse breakdown voltage in excess of −40 V. At −30 V, the measured reverse saturation current is 1.2 nA/μm^2, and falls to 89 pA/μm^2 at −1 V. This is comparable to a reported leakage current density of 70 pA/μm^2 for a 800 Å SiGeC p-i-n photodetector [16] and about 100 pA/μm^2 at −1 V for a Ge p-i-n photodetector on Si [17] which used $a \geq 2$ μm buffer layer.

For a low C concentration in the Ge$_{1-x}$C$_x$ epilayer, the conduction-band discontinuity is estimated as 0.05 eV, obtained from the electron affinity of Si (4.05 eV) and of pure Ge (4.0 eV) [18], while the valence-band discontinuity is 0.51 eV. Alternatively, the band offsets of SiGe heterojunctions have been well studied and predict a valence-band offset of 0.68 eV [19] for a pseudomorphic Si/Ge interface. However, the band discontinuities are well defined only if the in-plane lattice parameter is continuous across the interface [19]. This is not the case here due to partial or complete relaxation of the epilayer. However, both models predict a type II heterojunction for Si/GeC with low C%.

Calculation from the proposed band diagram indicates that the built-in voltage should be about 0.37 eV. The discrepancy between the theoretical and the measured turn-on voltage of 0.15–0.2 eV could be caused by 1) poor crystalline quality of the GeC/Si interface which has a large number of dislocations, 2) bandgap narrowing in the GeC epilayer due to residual strain at the GeC/Si heterojunction interface, or 3) bandgap narrowing due to highly doped GeC. The dislocations can act as trapping centers and allow defect-assisted tunneling of electrons from the conduction band of Si to these trapping centers within the bandgap of GeC and subsequent emission of electrons from these states to the conduction band of GeC. Therefore, the initiation of the turn-on voltage is obscured and leads to the slight s-shape in the $I-V$ curve at low bias.

Photoresponsivity was observed for the GeC/Si heterojunction diodes. Lasers with wavelengths of 1.3 μm (hν >

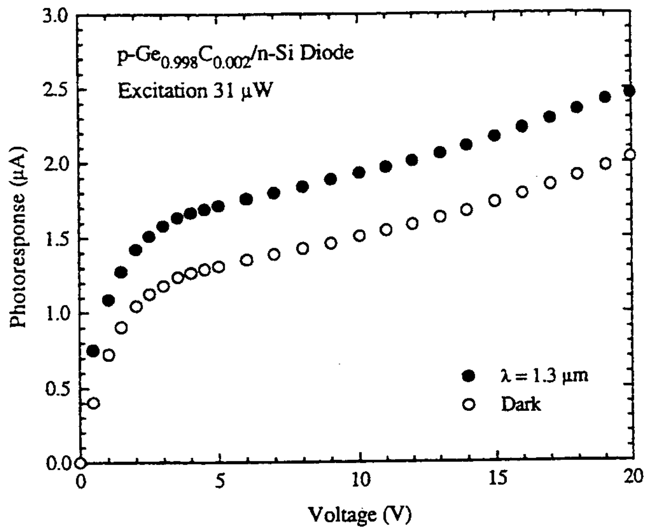


Fig. 3. The dark current (o) and photoresponse (•) of the reverse-biased p-$Ge_{0.998}C_{0.002}$/n-Si diode under laser excitation($\lambda = 1.3$ μm) of photon energy above the GeC bandgap but below the bulk Si bandgap.

GeC only) and 0.78 μm (hν > GeC and Si) were used to measure photoresponsivity. An appreciable photoresponse was measured using 0.78 μm laser excitation. This is to be expected if photons are absorbed in the surrounding bulk Si, around the mesa perimeter, and then photo-induced carriers diffuse to the p-n heterojunction. However, photoresponsivity was also observed when excitation was switched to below the Si bandgap ($\lambda = 1.3$ μm). The small laser spot was able to be focused between the periphery of the mesa and the upper metal contact. A photocurrent of 0.45 μA was measured, as shown in Fig. 3, which corresponds to a responsivity of 1.45×10^{-2} A/W and an external quantum efficiency (EQE) of 1.4%. These results are significant because the diode did not have an anti-reflection (AR) coating and the light collection region is confined to a narrow region (tens of angstroms) at the p^+-GeC/n^--Si junction. This leads to a much shorter path length within the p^+-GeC depletion region than a previous normal incidence 800 Å SiGeC p-i-n photodetector which reported EQE ~1% at 1.3 μm [16].

IV. CONCLUSION

In conclusion, the first p-n diode made from a heterojunction of epitaxial p-type $Ge_{0.998}C_{0.002}$ on an n-type Si substrate was demonstrated. The $I-V$ curve of the p-GeC/n-Si diode indicated reasonable reverse saturation current (89 pA/μm^2 at -1 V) and high reverse breakdown voltage in excess of -40 V. Some evidence in the $I-V$ curves suggests a small tunnel current was combined with a large forward biased diffusion current. Despite the large number of dislocations and defects at the heterojunction, a photoresponse from the $Ge_{0.998}C_{0.002}$ epilayer was observed by 1.3-μm laser excitation with a measured EQE of 1.4%.

ACKNOWLEDGMENT

The authors wish to thank F. Chen for Hall measurements.

REFERENCES

[1] R. A. Soref, "Silicon-based optoelectronics," in *Proc. IEEE*, 1993, vol. 81, pp. 1687–1706.
[2] R. People, "Physics and applications of Ge_xSi_{1-x}/Si strained layer heterostructures," *IEEE J. Quantum Electron.*, vol. QE-22, pp. 1696–1710, 1986.
[3] G. L. Patton, J. H. Comfort, B. S. Meyerson, E. F. Crabbé, G. J. Scilla, E. De Frésart, J. M. C. Stork, J. Y.-C. Sun, D. L. Harame, and J. N. Burghartz, "75-GHz f_T SiGe-base heterojunction bipolar transistors," *IEEE Electron Device Lett.*, vol. 11, pp. 171–173, 1990.
[4] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGouos, "Growth and strain compensation effects in the ternary $Si_{1-x-y}Ge_xC_y$ alloy system," *Appl. Phys. Lett.*, vol. 60, pp. 3033–3035, 1992.
[5] B. Dietrich, H. J. Osten, H. Rücker, M. Methfessel, and P. Zaumseil, "Latrice distortion in a strain-compensated $Si_{1-x-y}Ge_xC_y$ layer on silicon," *Phys. Rev. B*, vol. 49, pp. 17185–17190, 1994.
[6] F. J. Guarin, S. S. Iyer, A. R. Powell, and B. A. Ek, "Growth and strain symmetrization of Si/Ge/C/Sn quaternary alloys by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 68, pp. 3608–3610, 1996.
[7] H. J. Osten and J. Klatt, "*In situ* monitoring of strain relaxation during antimony-mediated growth of Ge and $Ge_{1-y}C_y$ layers on Si(001) using reflection high energy electron diffraction," *Appl. Phys. Lett.*, vol. 65, pp. 630–632, 1994.
[8] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, "Growth of germanium-carbon alloys on silicon substrate by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 67, pp. 1865–1867, 1995.
[9] L. D. Lanzerotti, A. St. Amour, C. W. Liu, J. C. Sturm, J. K. Wantanabe, and N. D. Theodore, "Si/$Si_{1-x-y}Ge_xC_y$/Si heterojunction bipolar transistors," *IEEE Electron Device Lett.*, vol. 17, pp. 334–337, 1996.
[10] A.-S. T. Khan, B. A. Orner, P. R. Berger, J. Kolodzey, F. J. Guarin, and S. S. Iyer, "Photoluminescence of $Ge_{1-x}C_x$ alloys grown on (100) Si substrates," presented at *1996 Spring Meet. Mat. Res. Soc.*, San Francisco, CA, Apr. 8–12.
[11] X. Shao, S. L. Rommel, B. A. Orner, P. R. Berger, and J. Kolodzey, "Low resistance ohmic contacts to p$Ge_{1-x}C_x$ on Si," *IEEE Electron Device Lett.*, vol. 18, pp. 7–9, 1997.
[12] R. I. Scace and G. A. Slack, "Solubility of carbon in silicon and germanium," *J. Chem., Phys.*, vol. 30, pp. 1551–1555, 1959.
[13] O. F. Sankey, A. A. Demkov, W. T. Petuskey, and P. F. McMillan, "Energetics and electronic structure of the hypothetical cubic zineblendo form of GeC," *Modeling Simul. Mater. Eng.*, vol. 1, pp. 741–754, 1993.
[14] M. Todd, J. Kouvetakis, and D. J. Smith, "Synthesis and characterization of heteroepitaxial diamond-structured $Ge_{1-x}C_x$ (x = 1.5–5.0%) alloys using chemical vapor deposition," *Appl. Phys. Lett.*, vol. 68, pp. 2407–2409, 1996.
[15] A.-S. T. Khan, "Photoluminescence and process issues of SiGeCSn alloys," Ph.D. dissertation, Univ. Delaware, Newark, 1996.
[16] F. Y. Huang and K. L. Wang, "Normal-incidence epitaxial SiGeC photodetector near 1.3-μm wavelength grown on Si substrate," *Appl. Phys. Lett.*, vol. 69, pp. 2330–2332, 1996.
[17] S. Luryi, A. Kastalsky, and J. C. Bean, "New infrared detector on a silicon chip," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 1135–1139, 1984.
[18] S. M. Sze, *Physics of Semiconductor Devices.* New York: Wiley, 1981, p. 850.
[19] C. C. Van de Walle, "SiGe heterojunctions and band offsets," *Properties of Strained and Relaxed Silicon Germanium, EMIS Data Reviews Series*, no. 12, p. 114, 1994.

Appendix 31

ELSEVIER

Thin Solid Films 321 (1998) 47–50

thin solid films

Electrical and optical properties of phosphorus doped $Ge_{1-y}C_y$

M.W. Dashiell[a,*], R.T. Troeger[a], K.J. Roe[a], A-S. Khan[a], B. Orner[a], J.O. Olowolafe[a], P.R. Berger[a], R.G. Wilson[b], J. Kolodzey[a]

[a] *Department of Electrical Engineering, University of Delaware, Newark, DE 19716, USA*
[b] *Hughes Research Laboratory, Malibu, CA 90265, USA*

Abstract

In situ n-type doping was investigated for $Ge_{1-y}C_y$/Si heteroepitaxial layers ($y \sim 0.001$) for a potential optoelectronic material compatible with Si. Using a solid GaP sublimation source for phosphorus doping, epitaxial $Ge_{1-y}C_y$ films were in situ doped on Si(100) substrates during solid source molecular beam epitaxy and we compare their electrical and optical properties with those of epitaxial Ge on Si. Infrared absorption revealed red shifts in the absorption of visible light with increasing P doping for both $Ge_{1-y}C_y$ and Ge. The index of refraction decreases for $Ge_{1-y}C_y$ layers compared with Ge. Free carrier absorption increased with increasing phosphorus concentrations, following a wavelength dependence of $\sim\lambda^{3.1}$ in the region of 10–20 μm for heavily doped material. Addition of C did not affect the incorporation of P donors in the grown layers or the electrical activation of the donors. An increase in the electron mobility for heteroepitaxial $Ge_{1-y}C_y$ layers compared with Ge was observed for the doping levels studied. © 1998 Elsevier Science S.A. All rights reserved

Keywords: Germanium carbon; Phosphorus; Doping; Optical; Electrical; Molecular beam epitaxy

---

1. Introduction

Binary and ternary alloys of Si, Ge and C are currently under investigation for use in heterostructure and opto-electronic devices [1,2]. Because of the low solubility of C in Si ($\sim 3 \times 10^{18}$ cm^{-3}) and Ge ($\sim 1 \times 10^{8}$ cm^{-3}) these alloys are typically grown far from equilibrium by molecular beam epitaxy (MBE) [3,4], chemical vapor deposition (CVD) [5] and solid phase epitaxy (SPE) [6]. Carbon concentrations far in excess of the equilibrium solubility have been reported for both $Si_{1-y}C_y$ [7] and $Ge_{1-y}C_y$ [8].

Significant development efforts are now concentrated on an understanding of C on the structural, optical and electrical properties of the $Si_{1-y}C_y$ and $Si_{1-x-y}Ge_xC_y$ alloys. Most of these efforts focus on the Si-rich alloy, due to the ability to strain compensate the layer, so that thick, dislocation free $Si_{1-x-y}Ge_xC_y$ layers may be grown on Si substrates [9]. The germanium rich spectrum of the alloy has not been as thoroughly investigated because of the large lattice mismatch between the epitaxial layer and silicon substrates. The large mismatch results in critical thicknesses of ~20 monolayers or less for Ge/Si epitaxy, depending on the growth parameters [10].

Despite the large lattice mismatch between Ge and Si, the development of group IV heterostructures using germanium-rich alloys opens a wide variety of possible device applications on silicon substrates, such as infrared photodetectors, modulation doped field effect transistors (MODFETs) and optical waveguides. Significant Ge fractions ($x \geq 0.50$) are necessary for efficient infrared detection of 1.3 and 1.55 μm light for fiber optic communication. Low dislocation density $Si_{1-x-y}Ge_xC_y$/Si pin photodiodes with $x = 0.60$ have been fabricated with external quantum efficiencies of 1% at 1.3 μm [11]. Higher Ge concentrations are expected to increase the quantum efficiency due to the corresponding decrease in bandgap. Adding small amounts of carbon to epitaxial germanium may reduce the total defect density associated with lattice mismatch, surface roughness and interfacial defects during the heteroepitaxy of germanium-rich alloys on silicon. Fukami [12] reported a 6-fold decrease in reverse leakage current with implantation of C atoms in SiGe/Si heterostructure diodes. The decrease in leakage current was attributed to improvement in crystalline quality due to C doping. Band structure modifications by alloying carbon with germanium may be possible as has been demonstrated for $Si_{1-y}C_y$ and $Si_{1-x-y}Ge_xC_y$ alloys

---

[*] Corresponding author. Tel.: +1 302 8318959; fax: +1 302 8314316; e-mail: dashiell@ee.udel.edu

0040-6090/98/$19.00 © 1998 Elsevier Science S.A. All rights reserved
PII S0040-6090(98)00441-6

[13]. Strained two-dimensional $Ge_{1-y}C_y$ layers with enhanced C solubility ($y = 0.01$) grown on Si(100) in the presence of surfactants have been demonstrated by MBE [10].

Vital to the success of any new semiconductor material is the ability to precisely control the carrier concentrations by introducing dopant atoms. We are unaware of any experimental study of n-type doping in $Ge_{1-y}C_y$ epitaxial films on Si(100). We report on the electrical and optical properties of P-doped (N_D~10^{18}–10^{20} cm^{-3}) $Ge_{1-y}C_y$ epitaxial layers grown by MBE on Si(100) substrates. High doping levels such as these are necessary for development of such devices as the $Ge_{1-y}C_y$ Esaki tunneling diode on Si. Of the group IV elements, Ge Esaki diodes exhibit the highest peak to valley current and packaged configurations exhibit switching speeds as high as 22 ps. Germanium Esaki diodes are not monolithically integrated into silicon based technology due to the large lattice mismatch between the two elements.

2. Experimental

In this paper we discuss the results of in situ phosphorus doping of epitaxial $Ge_{1-y}C_y$ films grown on Si(100) substrates. Samples were grown on 7.6-cm Si(100) substrates in an EPI-620 MBE chamber [3]. Pure Ge samples were epitaxially grown under equivalent conditions for comparison. All samples were grown at a substrate temperature of 400°C. Sharp streaks on the reflection high energy electron diffraction (RHEED) screen indicated a flat, single crystalline surface.

Zone refined germanium was thermally evaporated, at 0.20 Å/s, from a pyrolytic boron nitride (pBN) crucible contained in a standard EPI effusion cell. Carbon was introduced into the molecular beam by passing current through a high purity, pyrolytic-graphite filament. Phosphorus doping was achieved by in situ evaporation of GaP from a pBN crucible at temperatures from 650 to 750°C with P_2 being the dominant species in the beam [14]. A pBN gettering module is employed to minimize residual gallium in the beam. The epitaxial layers were 1μm in thickness. Previous work has shown that thick Ge on Si grown at 400°C demonstrates a low density of defects at the surface, determined by etch pit densities and spectroscopic ellipsometry [15].

Secondary ion mass spectrometry (SIMS) using a CAMECA 4f-sector magnet (Charles Evans and Associates) measured the phosphorus content and many other residual elements associated with the sources and their respective crucibles. A 14.5-KeV Cs primary beam was used for Si, Ge and P and an 8.0-KeV O_2 beam was used to measure the B and Ga concentrations. Phosphorus levels varied from 10^{18} to 10^{20} cm^{-3}. Concentrations of boron from the pBN crucible were less than 10^{16} cm^{-3}, the detection limit of the measurement. Due to insufficient gettering of Ga from the GaP sublimation cell, a significant concentration of Ga was incorporated into the epitaxial layer (between 1×10^{16} cm^{-3} and 1×10^{17} cm^{-3}), similar to previous studies for SiGe MBE [16]. Carbon concentrations were estimated from X-ray rocking curves of the (004) symmetric and (224) and (511) asymmetric reflections. Rutherford backscattering spectrometry using a 4.23-MeV $^4He^{++}$ ion beam detected C signal near the detection limit of the system (≤0.003). The C concentration estimated by the relaxed lattice parameter of the $Ge_{1-y}C_y$ layer is y~0.001. Room temperature infrared absorption spectroscopy did not reveal an absorption mode attributed to substitutional C in Ge. This mode was observed by Hoffman [8] for y~0.007 using low temperature (9 K) infrared spectroscopy. The small C concentration and thickness for our samples may however, produce vibrational signal below the room temperature sensitivity of our instrument.

Optical transmission of the epitaxial layers was measured with a Nicolet 740 Fourier transform infrared spectrometer (FTIR). Absorption due to the Si substrate was corrected by subtracting the absorption spectra of a reference Si substrate. Hall effect measurements of P-doped epitaxial $Ge_{1-y}C_y$ and Ge layers were carried out using a standard van der Pauw structure under a magnetic field of 3.3 kG. Measurements were taken for two polarities of magnetic fields to eliminate the misalignment voltage of the Hall contacts.

3. Results

3.1. Carrier concentration and Hall effect

SIMS measurements revealed that phosphorus concentrations ranged from 10^{18} to 10^{20} cm^{-3}. Phosphorus concentrations for epitaxial $Ge_{1-y}C_y$ and Ge layers did not differ for equivalent P cell temperatures. This indicates that small C fractions do not affect the incorporation of P into the grown layer. Electron concentrations measured by the Hall effect correlated well with SIMS, indicating the complete activation of P donors for both $Ge_{1-y}C_y$ and Ge epitaxial layers. We can conclude that P incorporates on substitutional sites during $Ge_{1-y}C_y$ and Ge epitaxial growth, i.e. for small amounts, C is electrically neutral in $Ge_{1-y}C_y$. Faschinger [17] demonstrated for $Si_{0.99}C_{0.01}$ alloys that the incorporation and activation of extrinsic donors (Sb) is very similar to that in pure Si. Their group also demonstrated the important result that electron mobility in strained $Si_{1-y}C_y$ with N_D~10^{16} cm^{-3} is not limited by scattering at lattice deformations induced by C, but rather by ionized impurity scattering.

Fig. 1 shows the room temperature Hall mobility for Ge and $Ge_{1-y}C_y$ layers for two different donor concentrations. An increase in the room temperature mobility is observed for heteroepitaxial $Ge_{1-y}C_y$ films compared to Ge films grown under equivalent conditions. X-ray diffraction of the epitaxial layers revealed an average decrease of 30% of the full width at half maximum (FWHM) for $Ge_{1-y}C_y$ compared with Ge suggesting a reduced defect density for

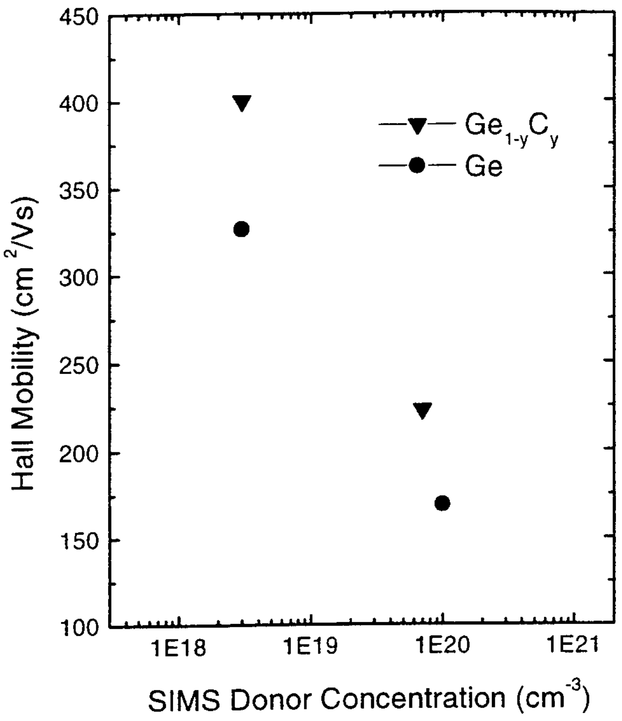


Fig. 1. Hall mobility of phosphorus doped Ge$_{1-y}$C$_y$ (▼) and Ge (●) epitaxial layers on Si(100) at 296 K.

Ge$_{1-y}$C$_y$ epitaxial layers. The findings of Fukami [12] and Shao [18] describing the reduced reverse leakage current and increased reverse breakdown voltages for Si$_{1-x-y}$Ge$_x$C$_y$ and Ge$_{1-y}$C$_y$ diodes with introduction of C support this observation.

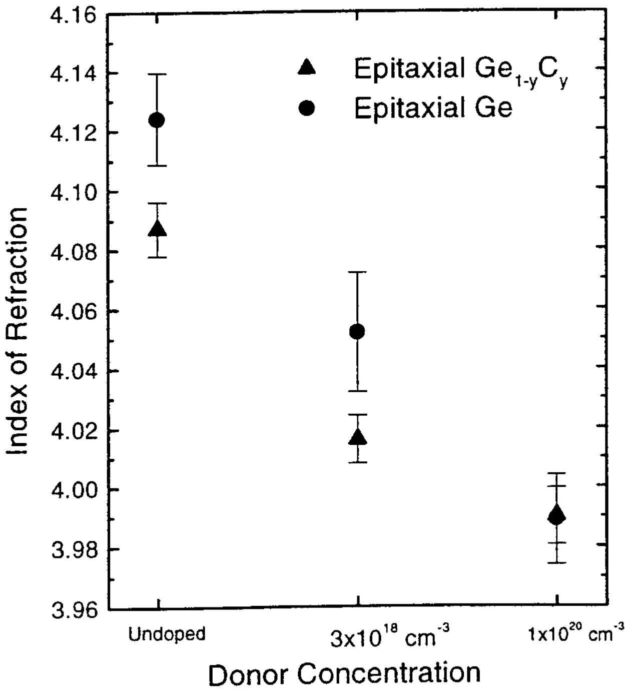


Fig. 2. Refractive index versus donor concentration for Ge$_{1-y}$C$_y$ (▲) and Ge (●) epitaxial films on Si(100). The refractive index was computed from the interference fringes of the transmission spectrum.

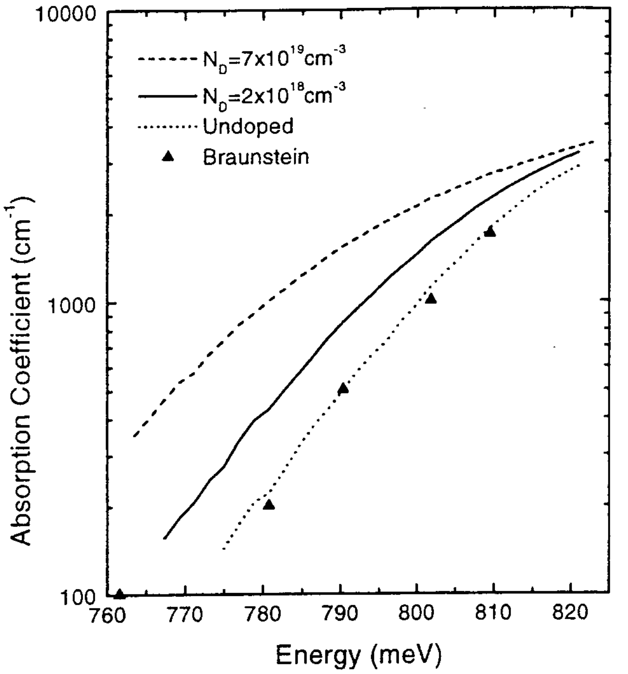


Fig. 3. Absorption coefficient versus photon energy of Ge$_{1-y}$C$_y$ layers on Si(100) for $N_D = 7 \times 10^{19}$ cm^{-3} (dashed line), $N_D = 2 \times 10^{18}$ cm^{-3} (straight line) and undoped (dotted line). Included are values for intrinsic bulk Ge [22].

3.2. Optical transmission

Transmission spectroscopy was performed in the visible and infrared regions to observe the near band edge absorption spectra and free carrier absorption of the epitaxial films. For energies below the bandgap the optical constants and thickness of the epitaxial layers were determined by curve fitting interference fringes resulting from multiple reflections at the air/film and film/substrate interface [19].

By fitting the measured transmission spectra of the Ge$_{1-y}$C$_y$ and Ge films in the region of 560–700 MeV we determined the thickness of the epitaxial film and the index of refraction at energies near the bandgap. Fig. 2 shows the index of refraction of the epitaxial Ge$_{1-x}$C$_x$ as a function of donor concentration and compares them to Ge epitaxial layers grown under identical conditions. Introducing carbon into epitaxial Ge films doped with P decreases the refractive index near the absorption edge.

Absorption coefficients above the band edge were computed from the measured transmission spectra and the optical constants as described in reference [20]. Fig. 3 illustrates the absorption coefficients (α) of phosphorus doped Ge$_{1-y}$C$_y$ films grown epitaxially on Si(100) for $\alpha > 100$ cm^{-1}. The absorption edge experiences a red shift with increasing phosphorus concentrations for both Ge$_{1-y}$C$_y$ and Ge films. Included in the graph is published data on high purity Ge [21]. Note that undoped Ge$_{1-y}$C$_y$ epitaxial layers exhibit the same absorption coefficient as does intrinsic bulk germanium for $\alpha > 100$ cm^{-1}, thus a significant band structure modification was not observed by optical absorption for these C concentrations. Because of the thickness of the epitaxial layers, absorption coefficients less than 100 cm^{-1} can not be accurately measured, thus the minimum indirect energy gap was not determined by this method. The absorption coefficient of Ge for $\alpha > 100$ cm^{-1} is expected to have transition contributions from the direct conduction band valley at Γ and from indirect valleys at L and X. Our results are consistent with those of Pankove [22] who reported a shrinkage of the energy gaps at $k = (111)$ and the direct valley at $k = (000)$ with heavy doping of bulk Ge.

3.3. Free carrier absorption

Absorption coefficients of phosphorus doped $Ge_{1-y}C_y$ and Ge epitaxial layers were calculated in the regions from 2.5 to 25 μm from transmission data. Free carrier absorption of heavily doped $Ge_{1-y}C_y$ and Ge layers follows a wavelength dependence of $\alpha \sim \lambda^{3.1}$ for $N_D \sim 10^{20}$ cm^{-3}. For lighter doping a weak wavelength dependence was observed for both materials. No significant change was observed with the addition of C. This behavior supports the observation that C is electrically neutral in Ge and does not change the dominant room temperature scattering mechanism.

3.4. Conclusions

Optical and electronic properties of epitaxial $Ge_{1-y}C_y$ layers on Si(100) Si were investigated and compared with epitaxial Ge on Si. SIMS and Hall effect analysis indicate that C does not affect the incorporation of P into the film and that C is electrically neutral in $Ge_{1-y}C_y$ for $y \sim 0.001$. Full activation of P donors is observed by close agreement between SIMS and electron concentration measurements. X-ray and high energy electron diffraction indicate a single crystalline structure. A decrease in the bulk lattice constant compared to pure Ge was observed with the addition of carbon ($y \sim 0.001$). A strong alloying affect due to C was not observed in the optical absorption of $Ge_{1-y}C_y$ for $\alpha \geq 100$ cm^{-1}, although a decrease in the index of refraction may indicate a change in the dielectric properties with addition of C. We conclude that small alloy concentrations of C during the heteroepitaxial growth of $Ge_{1-y}C_y$ on Si have enhanced the electronic properties of epitaxial $Ge_{1-y}C_y$. Initial results indicate that the enhancement may be due to a change in the structural properties of the heteroepitaxial layer due to carbon, rather than a modification of the band structure. These results indicate that $Ge_{1-y}C_y$/Si heteroepitaxy is an interesting new material system that may be applicable for optoelectronic device applications, particularly for infrared detection at 1.3 and 1.55 μm.

References

[1] R.A. Soref, Proc. IEEE 81 (1993) 1687.
[2] A.R. Powell, K. Eberl, F.E. LeGoues, B.A. Ek, S.S. Iyer, J. Vac. Sci. Technol. B 11 (1993) 1064.
[3] J. Kolodzey, P.A. O'Neil, S. Zhang, B. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite, S.I. Shah, Appl. Phys. Lett. 67 (1995) 1865.
[4] H.J. Osten, E. Bugiel, P. Zaumseil, Appl. Phys. Lett. 64 (1994) 3440.
[5] A. St. Amour, C.W. Liu, J.C. Sturm, Y Lacroix, M.L.W. Thewolt, Appl. Phys. Lett. 67 (1995) 3915.
[6] J.W. Strane, H.J. Stein, S.R. Lee, B.L. Doyle, S.T. Picraux, J.W. Mayer, Appl. Phys. Lett. 63 (1993) 2786.
[7] H.J. Osten, M. Kim, K. Pressel, P. Zaumseil, J. Appl. Phys. 80 (1996) 6711.
[8] L. Hoffman, J.C. Bach, B.B. Nielson, P. Leary, R. Jones, S. Oberg, Phys. Rev. B 55 (1997) 11167.
[9] K. Eberl, S.S. Iyer, S. Zollner, J.C. Tsang, F.K. LeGoues, Appl. Phys. Lett. 60 (1992) 3035.
[10] H.J. Osten, J. Klatt, Appl. Phys. Lett. 65 (1994) 630.
[11] F.Y. Huang, K.L. Wang, Appl. Phys. Lett. 69 (1996) 2330.
[12] A. Fukami, K. Shoji, T. Nagano, C.Y. Yang, Appl. Phys. Lett. 57 (1990) 2345.
[13] K. Brunner, W. Winter, K. Eberl, N.Y. Jin Phillip, F. Phillipp, J. Cryst. Growth 175/176 (1997) 451.
[14] T. Shitara, K. Eberl, Appl. Phys. Lett. 65 (1994) 356.
[15] K.E. Junge, R. Lange, J.M. Dollan, S. Zollner, M. Dashiell, B.A. Orner, J. Kolodzey, Appl. Phys. Lett. 69 (1996) 4084.
[16] G. Lippert, H.J. Osten, D. Kruger, J. Cryst. Growth 157 (1995) 304.
[17] W. Faschinger, S. Zerlauth, G. Bauer, L. Palmetshufer, Appl. Phys. Lett. 67 (1995) 3922.
[18] X. Shao, S.L. Rommel, B.A. Orner, H. Feng, M. Dashiell, R.T. Troeger, J. Kolodzey, P.R. Berger, Appl. Phys. Lett. 72 (1997) in press.
[19] O.S. Heavens, Optical Properties of Thin Solid Films, Dover, New York, 1991.
[20] B.A. Orner, F. Chen, D. Hits, M.W. Dashiell, J. Kolodzey, Proc. SPIE 3007 (1997) 152.
[21] R. Braunstein, A. Moore, F. Herman, Phys. Rev. 109 (1958) 695.
[22] J.I. Pankove, in A.F. Gibson, R.E. Burgess (eds.), Progress in Semiconductors, Vol. 9, Heywood, London, 1965, p. 48.

PHOTOLUMINESCENCE OF SiSnC ALLOYS GROWN ON (100) Si SUBSTRATES

N. Wright [*], A.T. Khan [*], P.R. Berger [*], F.J. Guarin [**], S.S. Iyer [**]
[*]Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716
[**]IBM Microelectronics Division, 1580 Route 52, Hopewell Junction, NY 12533

ABSTRACT

Photoluminescence (PL) of SiSnC alloys grown on Si (100) substrate by molecular beam epitaxy (MBE) has been investigated. The following epitaxial layers were investigated. The samples were similar in structure, and consisted of a 200 Å Si buffer layer grown on a (100) Si substrate followed by the respective alloy layer; a 4500 Å $Si_{0.955}Sn_{0.03}C_{0.015}$, a 1500 Å $Si_{0.96}Sn_{0.04}$, or a 1500 Å $Si_{0.985}C_{0.015}$. The layer composition was measured by Rutherford back scattering spectrometry (RBS) and confirmed by x-ray diffraction analysis (XRD). X-ray diffraction measurements of the layers confirmed the Sn and C were substitutional and the layers were pseudomorphic and coherently strained. The $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy layer was found to be strain compensated. PL spectra of all the layers revealed band edge luminescence as well as a very sharp peak at 0.767 eV superimposed on a very broad peak that exhibited excitonic behavior. The addition of C to the alloys resulted in a reduction of the bandgap, contrary to what is predicted by Vegard's Law. However, the addition of Sn results in a reduction in the bandgap, which was attributed to the bulk alloy effect and residual strain. The luminescence feature at 0.767 eV was found to be much more intense in alloys that contain carbon. We have observed quite similar deep-level well resolved PL spectra for SiGe, SiC, SiSn, SiGeC, SiSnC, and SiGeSnC alloy layers grown by MBE on Si (100) substrates. Previous studies on PL of silicon have reported a peak at 0.767 eV to be associated with oxygen (P line) and the broad peak to carbon/oxygen complexes.

INTRODUCTION

The growth of alloys of group IV elements has been under investigation recently as potential material systems for Si-based heterostructure device applications [1]. In the past decade extensive effort has been directed toward the SiGe material system, which has led to the recent commercial availability of a few SiGe devices and circuits. There has been some interest recently in the growth of Sn-based and C-based group IV semiconductor alloys. The carbon based alloys of SiC and SiGeC have received the most attention. There have also been a few investigations directed toward the Sn-based alloys. High quality pseudomorphic crystalline layers of Sn and C based alloys have been grown by a few groups. The Sn-based alloys have been grown by MBE [2], and C-based alloys have been grown by chemical vapor deposition (CVD) [3] and MBE [4]. These studies indicate that only alloys containing a few percent of C or Sn can be grown pseudomorphically on a Si substrate. This is due mainly to the following factors. The alloys suffer from metastability problems, large lattice and bond length mismatches, and low Sn and C solid solubilities.

Alloy layers that exceed a few percent of C were found to be plagued by defects, non-planar growth, or SiC precipitates. This can be directly attributed to the extremely low solubility of C in Si ($\sim 3 \times 10^{18}$ cm^{-3}) [5], and the large difference in the Si and C bond lengths. Si has a bond length of 2.35 Å, diamond a bond length of 1.55 Å, and zinc-blende β-SiC phase has a bond length of 1.89 Å [6]. The bond length mismatch of about 40 % causes large local strain and according to the Si-C phase diagram, stoichiometric SiC is the only stable compound. Previous work on the Sn-based alloys indicated that, it is very difficult to stabilize Sn in the Si alloy lattice. This is mainly due to the large difference in the Si and Sn lattice parameters. Grey tin has a lattice parameter of 6.489 Å, and

Mat. Res. Soc. Symp. Proc. Vol. 533 © 1998 Materials Research Society

Si has a lattice parameter of 5.431 Å, resulting in a 19.48 % lattice mismatch. Grey tin (α-Sn) crystallizes in the diamond structure, and at 13.2 °C it transforms into the tetragonal structure of metallic white tin (β-Sn). Sn has a very low solid solubility in crystalline silicon (~5 x 10^{19} cm^{-3}) [5]. Therefore the substitutional incorporation of C and/or Sn above the solid solubility limit requires far from thermodynamic equilibrium growth conditions, such as MBE. The incorporation of isoelectronic Sn or C into silicon will potentially offer new possibilities, for bandgap engineering, silicon compatible wide or narrow bandgap material, and large band offsets. This could lead to changes in the optical and electrical properties of the material. According to recent theoretical calculations [1], $Si_{1-x-y}Sn_xC_y$ alloys would ideally span the energy range of 0.08 eV to 5.48 eV. These alloys offer the possibility of tuning the bandgap energy over an enormous range by varying the compositions of their respective constituents. Additionally, the bandgap was predicted [7] to be direct for certain compositions of SiSn and SnC. They also have the potential for exact lattice matching to Si. Strained layers of SiSnC alloy semiconductors would find numerous applications in electronic and optoelectronic heterostructures. These alloys also offer the potential for high electron and hole mobilities. To the best of our knowledge the optical luminescence properties of pseudomorphic $Si_{1-x-y}Sn_xC_y$ alloys grown on Si substrates have not been reported in any previous studies.

EXPERIMENT

The details of the solid source MBE system used for the growth of the SiSnC layers are described elsewhere [2]. A thin 200 Å Si buffer layer was grown initially on a (100) Si substrate followed by the respective alloy layer. The $Si_{0.985}C_{0.015}$ and the $Si_{0.96}Sn_{0.04}$ layer was grown to a thickness of 1500 Å, and the $Si_{0.955}Sn_{0.03}C_{0.015}$ layer was grown to a thickness 4500 Å. The growth conditions were optimized to maximize the incorporation of C and/or Sn on Si lattice sites. The layer compositions were measured by Rutherford backscattering spectrometry (RBS) and x-ray diffraction analysis (XRD), which confirmed the Sn and C were substitutional and the layers were pseudomorphically strained. XRD of the SiSnC alloy layer showed it to be strain compensated and lattice matched to Si, the details are described elsewhere [2].

Photoluminescence spectra was recorded in standard lock-in configuration, using a dispersive 1-m high-resolution Jarell-Ash (Czerny-Turner) monochromator and detected by a liquid-nitrogen cooled Ge p-i-n photodetector (North Coast EO-817L). The samples were mounted on a cold finger in a temperature variable helium-flow cryostat. The excitation was provided by a multi-line cw Argon laser (488-514 nm) focused to a sample area of approximately 2 mm^2, with pump intensities between 0.3 and 5 W/cm^2. Data collection, and lock-in amplification were controlled by a desktop computer.

RESULTS

Low temperature PL spectra at 6 K for the $Si_{0.955}Sn_{0.03}C_{0.015}$ sample is presented in Fig. 1. The luminescence consists of a very intense deep-level luminescence and a few edge peaks superimposed on a broad peak. The most intense peak at 1.040 eV is ascribed to a no-phonon (NP) transition, the peak located at 1.022 eV is red-shifted by 18 meV and can be identified with the momentum-conserving (MC) transverse acoustic (TA) phonon replica in silicon, the peak located at 0.982 eV is red-shifted by 58 meV and corresponds to its transverse optic (TO) Si-Si phonon replica. Additional peaks were observed at 1.006 eV, 0.966 eV, and 0.933 eV. These peaks were red-shifted from the ascribed NP line by 34 meV, 70 meV, and 107 meV respectively. The PL spectrum for the $Si_{0.96}Sn_{0.04}$ and $Si_{0.985}C_{0.015}$ samples is presented in Fig. 2. The luminescence from the $Si_{0.96}Sn_{0.04}$

sample consists of a low intensity deep-level luminescence and two edge peaks. The most intense peak at 1.055 eV is ascribed to a NP transition whereas the second peak at 0.994 eV is red-shifted by 61 meV and corresponds to its transverse optic (TO) Si-Si phonon replica. This is to our knowledge the first observation of luminescence from a SiSn alloy, and the details are reported elsewhere [8]. The luminescence from the $Si_{0.985}C_{0.015}$ sample consists of an intense deep-level luminescence and two edge peaks. The most intense peak at 1.060 eV is ascribed to a no-phonon (NP) transition whereas the second peak at 1.002 eV is red-shifted by about 58 meV and corresponds to its transverse optic (TO) Si-Si phonon replica, for more details refer to the following paper [9]. The luminescence peak at 1.1089 eV was observed in all the samples and is attributed to a Si TO phonon transition within the Si substrate. Also, a deep-level broad band luminescence is observed in all the samples around 0.778 eV. The deep-level feature was much more intense in the alloys that contained carbon, and consisted of a very sharp peak located at 0.767 eV superimposed on a broad peak. Previous studies on Si have reported a sharp peak located at 0.767 eV to be associated with oxygen (P line) [10]. Similar deep-level broad band luminescence has been observed in $Si_{1-x}Ge_x$, $Si_{1-x-y}Ge_xC_y$, and $Si_{1-x}C_x$ layers grown by MBE, and were attributed to localized excitons in a strain field created by Ge platelets [11], the emission of deep pseudo-acceptors [12], or to carbon-oxygen complexes [3], respectively.

Soref [7], using a linear interpolation between the constituent elements, predicted the addition of Sn to Si would result in a red shift of the bandgap, and the addition of carbon a blue shift in the bandgap. The bandgap of all the alloys investigated were found to be red shifted. The $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy NP transition is shifted 113 meV below that observed in pure Si [13], and is additionally

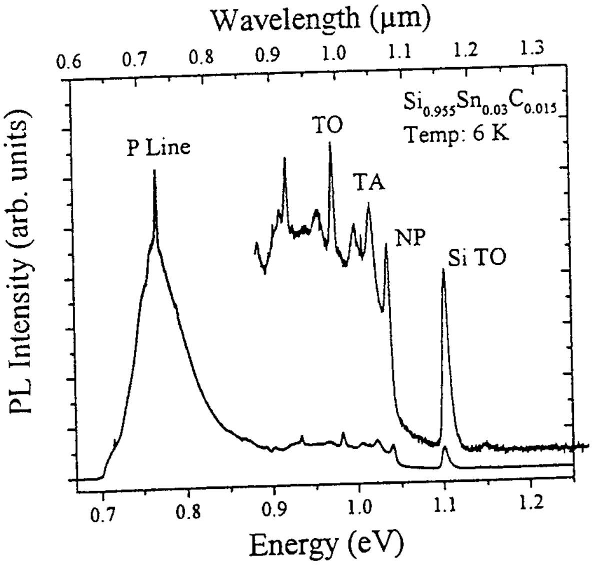


Fig. 1. PL spectra at 6 K of a strain compensated $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy layer grown by MBE on (100) Si substrate, showing band-edge and deep-level luminescence. Inset shows an expanded view of the spectra between the energy range 0.9 - 1.25 eV.

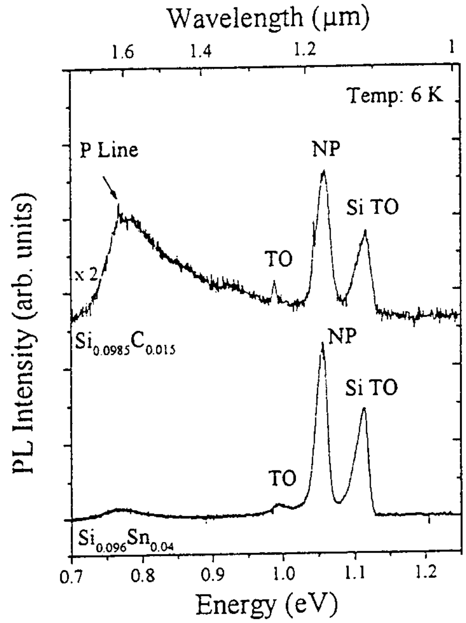


Fig. 2. PL spectrum of a $Si_{0.96}Sn_{0.04}$ and a $Si_{0.985}C_{0.015}$ alloy layer showing bandedge and deep-level luminescence.

shifted about 146 meV below an expected 1.186 eV luminescence position using Vegard's Law. XRD results [2] revealed a strain compensated and lattice matched $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy layer. We attribute this red shift to the lattice distortion within the alloy due to highly localized strain around the carbon and Sn atoms. The NP transition of the $Si_{0.96}Sn_{0.04}$ alloy is shifted 98 meV below that observed in pure Si [13]. We attribute this red shift to the reduction in band gap of the binary SiSn alloy. However, the observed NP transition is also shifted 55 meV below an expected 1.110 eV luminescence position using Vegard's Law for a SiSn alloy composition with 4% Sn. This additional red shift we attribute to reduction in the bandgap caused by residual strain in the pseudomorphic layer and highly localized strain around the Sn atoms. The NP transition of the $Si_{0.985}C_{0.015}$ alloy is shifted 93 meV below that observed in pure Si [13], and is additionally shifted about 159 meV below an expected 1.2182 eV luminescence position using Vegard's Law for a $Si_{1-x}C_x$ alloy composition with 1.5 % C. We attribute this red shift to both the lattice distortion within the alloy due to highly localized strain around the carbon atoms, and residual strain between the pseudomorphic alloy layer and the substrate. The lowering of the bandgap due to the pseudomorphic strain energy in the layer is not enough to explain the large red shift alone. The bandgap reduction in the carbon based alloys seems to follow the trend suggested by recent theoretical investigations [14] on the bandgap of $Si_{1-x}C_x$ alloys. They indicate that the addition of small carbon concentrations reduces the $Si_{1-x}C_x$ bandgap below Si, it turns semi-metallic around 10% C, and then increases above Si beyond 10% carbon.

The laser power dependence of the designated NP lines and the deep-level feature of the alloys were investigated. The NP peaks of the $Si_{0.96}Sn_{0.04}$ and the $Si_{0.985}C_{0.015}$ alloys reveal a linear increase in intensity with increasing laser power, while the $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy reveal a near linear dependence as is shown in Fig. 3. PL laser power dependence of this nature is characteristic of free exciton (FE) recombination. The deep-level feature, the line designated as the P line and the broad peak, exhibit a sub-linear almost square-root dependence, which indicates that a recombination center may be involved in the recombination process and have previously been attributed to carbon-oxygen complexes [3].

Temperature dependence of the designated NP peaks in Figs. 1 and 2 were also investigated. It was observed that the luminescence of the $Si_{0.96}Sn_{0.04}$ alloy layer persisted past 50 K, and the $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy layer luminescence persisted up to 100 K, but was degraded in intensity. The luminescence of the $Si_{0.985}C_{0.015}$ alloy layer however persisted past 200 K but was degraded in intensity. A further temperature increase resulted in the labeled NP and Si TO peaks merging into a single broad peak which persisted up to room temperature. As the temperature was increased, the designated NP peaks of the alloys all showed a decrease in PL intensity and a broadening of the linewidth in the direction of higher energy. This type of linewidth behavior is characteristic of the Maxwell-Boltzman distribution. PL temperature dependence of this nature is characteristic of free exciton (FE) recombination. Therefore, we attribute the labeled NP peaks as due to no phonon free exciton recombination. The full width at half maximum of the NP peak at 6 K of the $Si_{0.96}Sn_{0.04}$ and the $Si_{0.985}C_{0.015}$ layer were 18 meV and 19 meV respectively, which is rather broad compared to the intrinsic FE thermal line width, but this could be attributed to statistical fluctuations in the atomic distributions of the alloy as discussed by Robbins et al [15] and the lattice distortion of the alloy due to the large bond length difference between Si, Sn, and C as discussed by Demkov et al [14].

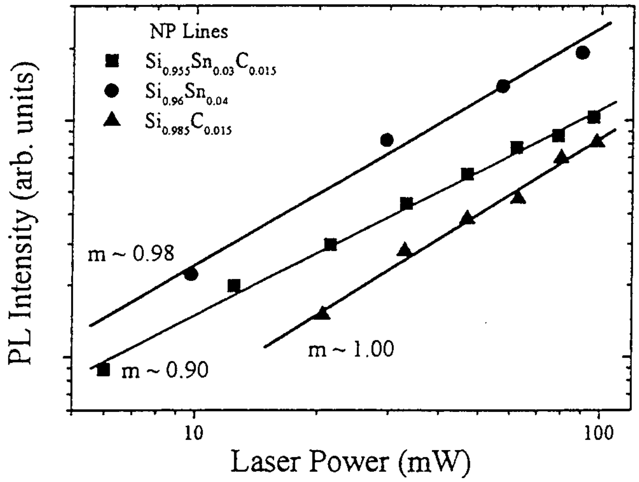


Fig. 3. PL laser power dependence of the assigned NP lines in the $Si_{0.96}Sn_{0.04}$, $Si_{0.985}C_{0.015}$, and $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy layers.

CONCLUSIONS

In conclusion, we have reported low temperature photoluminescence of a strain compensated $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy layer, a compressively strained pseudomorphic $Si_{0.96}Sn_{0.04}$ alloy layer and a tensially strained pseudomorphic $Si_{0.985}C_{0.015}$ alloy layer grown on Si (100) substrates by solid source MBE. Two general features have been observed, a deep-level broad band luminescence and band-edge luminescence consisting of a NP and TO replica in the $Si_{0.96}Sn_{0.04}$ and $Si_{0.985}C_{0.015}$ alloy layers, and a NP, TA and TO replicas in the $Si_{0.955}Sn_{0.03}C_{0.015}$ alloy layer. The deep-level luminescence of the alloys containing carbon was very intense and consisted of a very sharp peak located at 0.767 eV which was attributed to oxygen (P line) superimposed on a very broad peak. Temperature dependent PL analysis of the band-edge feature of the alloy layers indicated an exponential temperature dependence, which is characteristic of free excitonic recombination. The band-edge feature of the $Si_{0.96}Sn_{0.04}$ and $Si_{0.985}C_{0.015}$ layers exhibited a linear power dependence, and the $Si_{0.955}Sn_{0.03}C_{0.015}$ layer exhibited a near linear power dependence, which is also characteristic of free exciton recombination. The deep level feature in all the layers showed a sub-linear almost square-root dependence, which indicates that a recombination center may be involved in the recombination process. The energy gap of all the alloy layers was red-shifted with respect to Si and was attributed to lattice distortion within the alloy and residual strain between the alloy layer and the substrate where applicable.

REFERENCES

1. R.A. Soref, Proc. IEEE 81, 1687 (1993)
2. F.J. Guarin, S.S. Iyer, B.A. Ek, and A.R. Powell, Appl. Phys Lett. 68, 3608 (1996)
3. P. Boucaud, C. Francis, A. Larré, F.H. Julien, J.-M. Lourtioz, D. Bouchier, S. Bodnar, and J.L. Regolini, Appl. Phys. Lett. 66, 70 (1995)
4. S.S. Iyer, K. Eberl, M.S. Goorsky, F.K. LeGoues, J.C. Tsang, and F. Cardone, Appl. Phys. Lett. **60**, 356 (1992)
5. A.G. Milnes, Deep Impurities in Semiconductors, Wiley, New York, (1973)
6. H.J. Osten, E. Bugiel, and P. Zamuseil, J. Cryst. Growth **142**, 322 (1994)
7. R.A. Soref, J. Appl. Phys **70**, 2470 (1991)
8. A.T. Khan, P.R. Berger, F.J. Guarin, and S.S. Iyer, Appl. Phys. Lett. **68**, 3105 (1996)
9. A.T. Khan, P.R. Berger, F.J. Guarin, and S.S. Iyer, Thin Solid Films **294**, 122 (1997)
10. J. Weber, and R. Sauer, Mat. Res. Soc. Symp. Proc., **14**, 165 (1983)
11. J.-P. Noël, N.L. Rowell, D.C. Houghton, and D.D. Perovic, Appl. Phys. Lett. **61**, 690 (1992)
12. J. Denzel, K. Thonke, J. Spitzer, R. Sauer, H. Kibbel, H.-J. Herzog, and E. Kasper, Thin Solid Films **222**, 89 (1992)
13. P.J. Dean, J.R. Haynes, and W.F. Flood, Phys. Rev. **161**, 711 (1967).
14. A.A. Demkov, and O.F. Sankey, Phys. Rev. B **48**, 2207 (1993)
15. D.J. Robbins, L.T. Canham, S.J. Barnett, A.D. Pitt, and P. Calcott, J. Appl. Phys. **71**, 1407 (1992)

APPLIED PHYSICS LETTERS Appendix 33 VOLUME 72, NUMBER 7 16 FEBRUARY 1998

Carbon incorporation in $Si_{1-y}C_y$ alloys grown by molecular beam epitaxy using a single silicon–graphite source

M. W. Dashiell,[a] L. V. Kulik,[b] D. Hits, and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

G. Watson
Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716

(Received 2 October 1997; accepted for publication 16 December 1997)

Pseudomorphic $Si_{1-y}C_y$ alloys on silicon (100) were grown by molecular beam epitaxy using a single effusion source of silicon contained in a graphite crucible, producing carbon concentrations of $y=0.008$. The behavior of carbon incorporation using this source was studied as a function of growth temperature using x-ray diffraction and infrared spectroscopy, and was compared to previous studies, where $Si_{1-y}C_y$ was grown from separate silicon and graphite sources. An increased energy barrier for the surface diffusion of carbon was observed using the single silicon–graphite source. An infrared absorption mode near 725 cm^{-1}, observed for growth temperatures up to 700 °C, was attributed to a transitional phase between the loss of substitutional carbon and the formation of silicon carbide precipitates. © *1998 American Institute of Physics.* [S0003-6951(98)03707-3]

Recently, the alloying of C with Si and $Si_{1-x}Ge_x$ has attracted attention because of its ability to control the strain associated with the lattice mismatch to Si.[1,2] Carbon in the $Si_{1-y}C_y$ and the $Si_{1-x-y}Ge_xC_y$ systems modifies the alloy band structure and the band offsets for Si based heterostructures.[3] An increase of the substitutional C concentration in Si, by nearly four orders of magnitude more than the equilibrium solubility, has been predicted to arise from an enhanced surface solubility during epitaxial growth.[4] This result has been demonstrated by molecular beam epitaxy (MBE), but only for a small "growth window" where conditions such as the substrate temperature and growth rate are optimal for substitutional C incorporation.[5] The most common method of producing a C flux during MBE growth of $Si_{1-y}C_y$ is by sublimating high purity graphite in the presence of a Si flux.[1,5] However, previous studies have shown that the evaporation of Si in a graphite container will yield a molecular beam predominantly of Si, Si_2C, and SiC_2 species due to the reaction between Si and graphite.[6] Equivalent species were also obtained by the evaporation of SiC.[6,7]

In this letter, we report the structural properties of MBE grown $Si_{1-y}C_y$ on (100) Si using a single effusion cell of silicon contained in a graphite crucible. The behavior of C incorporation in the lattice for different substrate growth temperatures was studied using infrared absorption spectroscopy and x-ray diffraction. The amount of substitutional C was observed to decrease with increasing substrate temperature. Infrared absorption showed that the loss of substitutional C is due to the formation of an intermediate Si–C phase and SiC precipitates. Compared to previous studies using separate Si and graphite sources,[5] alloys grown with a single silicon–graphite source had an increased barrier to diffusion of C at the surface. We attributed this change to the difference in the C containing species of the molecular beam.

We grew 150 nm thick $Si_{1-y}C_y$ layers by MBE on the 2×1 reconstructed surface of Si (100) at substrate temperatures ranging from 400 to 750 °C, holding all other growth parameters constant. The MBE system was a model 620 manufactured by EPI Corporation, with a base pressure of less than 10^{-11} Torr.[8] Three in. diam float zone Si substrates were prepared by chemical degreasing and oxidation of the surface, followed by removal of the oxide in dilute hydrofluoric acid. Prior to growth, samples were heated to 200 °C for 45 min, then to 250 °C for 15 min, and finally the substrate was ramped to the growth temperature. Reflection high energy electron diffraction (RHEED) confirmed the 2×1 reconstruction of the Si surface prior to growth. X-ray diffraction was measured using a Philips Xpert-x-ray diffraction (XRD) diffractometer, using Cu $k\alpha_1$and $k\alpha_2$ radiation. Infrared absorption was measured with a Nicolet-740 Fourier transform infrared spectrometer.

The silicon–graphite effusion source was constructed from a high-temperature effusion cell manufactured by EPI. High purity Si pieces were heated inside a pyrolytic graphite crucible to produce the reaction forming the molecular beam. This resulted in a growth rate of approximately 0.33 A°/min at an effusion cell temperature of 1400 °C. At this temperature, the graphite crucible itself will have negligible vapor pressure for any C containing species.[9] Only the Si which reacted with the graphite crucible will have a significant vapor pressure, resulting in a molecular beam consisting of Si and C bonded with Si.

Figure 1 shows the x-ray diffraction and the infrared absorption spectra of $Si_{1-y}C_y$ layers grown by evaporation from the single silicon–graphite source. The $Si_{1-y}C_y$ alloy grown at 400 °C was amorphous, but was re-crystallized prior to both measurements by annealing at 750 °C for 20 min in N_2 gas, resulting in practically all C atoms occupying substitutional sites.[10] For samples grown at higher tempera-

---

[a] Electronic mail: dashiell@ee.udel.edu

[b] On leave from the Institute of Solid State Physics of the Russian Academy of Sciences, Chernogolovka, Moscow Region, Russia.

© 1998 American Institute of Physics

Appendix 33

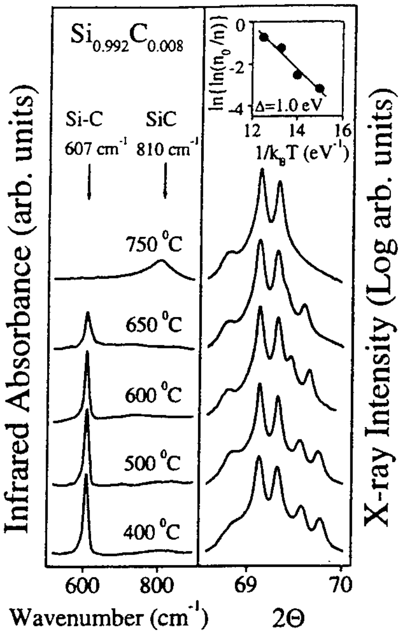


FIG. 1. Infrared absorption spectra and (004) x-ray reflections for pseudomorphic $Si_{1-y}C_y$ alloys grown on (100) silicon for substrate temperatures of 400–750 °C. The inset plots the substitutional carbon concentration normalized to the total carbon concentration vs the reciprocal growth temperature.

tures, annealing produced no significant change in infrared absorption or x-ray diffraction. The percentage of substitutional carbon in the layers was co-determined by the angular shift between the $k\alpha_2$ (004) reflections of silicon and $Si_{1-y}C_y$, and by the absorbance of the 607 cm^{-1} local vibrational mode (LVM) of C in Si. The C content of $y=0.008$ for the layer grown at 400 °C was measured after re-crystallization. X-ray rocking curve analysis of the symmetric (004) and asymmetric (224) reflections revealed that the layers were fully strained. X-ray diffraction and Raman measurements indicated that the C concentration was uniform across the wafer.

Infrared absorption measurements indicated that substitutional C exists in the layers for growth temperatures up to 700 °C. At a growth temperature of 750 °C, the LVM of C disappeared and the absorption peak at 810 cm^{-1} indicated that C had transformed to a SiC precipitate phase.[11] For intermediate growth temperatures (400–700 °C) the (004) x-ray diffraction peak of the $Si_{1-y}C_y$ layer shifted to smaller angles relative to the substrate peak with increased growth temperature (Fig. 1). The LVM of C in Si at 607 cm^{-1} experienced a decrease in intensity as the growth temperature was increased. These two measurements indicate a decrease in substitutional C versus increasing substrate temperature. An absorption peak near 725 cm^{-1} appeared for the intermediate temperatures along with a small SiC characteristic peak at 810 cm^{-1}. Figure 2 shows the infrared absorption spectra of $Si_{1-y}C_y$ alloys grown at 650 and 400 °C, where the infrared spectra of the layer grown at 400 °C was measured prior to recrystallization. Three distinct absorption features for the $Si_{1-y}C_y$ layer grown at 650 °C are the 607 cm^{-1} LVM of C, a small 810 cm^{-1} SiC precipitate peak, and the broad peak near 725 cm^{-1} resembling that of the as-grown amorphous $Si_{1-y}C_y$. The 725 cm^{-1} peak increased in intensity with growth temperature up to 700 °C, but disappeared for a growth temperature of 750 °C. At 750 °C, only the 810 cm^{-1} SiC peak remained. Our observation of the 725 cm^{-1} peak is consistent with a Si–C transitional phase, intermediate between the loss of substitutional C sites on the silicon lattice and the formation of SiC precipitates. The disordered nature of this transitional phase is supported by the results of Kimura *et al.*[12] who observed an absorption peak near 725 cm^{-1} for Si implanted with C ions which existed for annealing temperatures up to 900 °C.

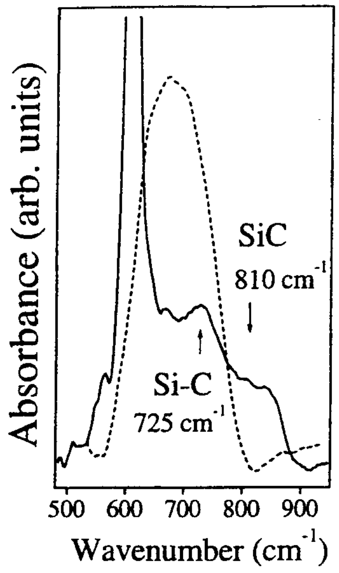


FIG. 2. The infrared absorption spectra of the as-grown amorphous $Si_{1-y}C_y$ alloy grown at 400 °C (dashed line) and crystalline $Si_{1-y}C_y$ grown at 650 °C (solid line). The alloy grown at 650 °C exhibits infrared peaks characteristic of the LVM of carbon in silicon, SiC precipitation, and an intermediate Si–C phase with an absorption peak centered at 725 cm^{-1}.

The behavior of C incorporation during MBE growth of $Si_{1-y}C_y$ can be explained by the diffusion and formation of Si–C interstitial defects near the surface.[5] The number of C atoms, n, on substitutional lattice sites can be modeled by:

$$n(T,r)=n_o(r,f)\exp\left(\frac{-ak}{4r}\right), \qquad (1)$$

where n_o is the total C concentration, r is the silicon growth rate, f is the carbon flux, a is the Si lattice constant, T is the growth temperature, and the exponential rate term

$$k=k_o\exp(-\Delta/k_BT) \qquad (2)$$

describes the transfer of C from substitutional sites to interstitial Si–C defects at the surface. The activation energy is designated by Δ and k_B is the Boltzmann constant. The term $a/4r$ in Eq. (1) is the time for 1 monolayer of material to be grown, after which C will be effectively "frozen in" the lattice and thus subject only to bulk diffusion.

The top right inset of Fig. 1 displays the behavior of $\ln[\ln(n_o/n)]$ vs $1/k_BT$ for $Si_{1-y}C_y$ layers grown by the evaporation of Si in a graphite crucible, where the reference n_o was taken to be the substitutional C concentration of the layer grown at 400 °C after re-crystallization. The best fit to this data gives the energy of formation of interstitials Si–C defects of $\Delta=1.0\pm0.2$ eV in the temperature range up to the threshold for SiC precipitation. The growth of $Si_{1-y}C_y$ layers on (100) Si by the sublimation of graphite from a heated filament separate from the pure silicon beam yields a smaller activation energy of $\Delta = 0.5$ eV.[5] The reason for the increased activation energy observed for $Si_{1-y}C_y$ grown from the single silicon–graphite source may be attributed to the chemical composition of the C containing species in the molecular beam. Evaporation from the single silicon–graphite source produces a molecular beam where the C containing molecular species is bonded to Si.[6] In contrast, evaporation from pure graphite does not.[9] We infer that the bonding of C to Si in the molecular beam tends to decrease surface diffusion of C, possibly by increasing the probability that the small C atom will be immediately buried under a Si atom. Once the C atom has been buried under 1 or more monolayers of Si, its position will be effectively "frozen in" and will be subject only to bulk diffusion.[4,5] Substitutional C incorporation has been shown to depend on the gas source used during rapid thermal chemical vapor deposition (RTCVD) of $Si_{1-x-y}Ge_xC_y$.[13] Mi *et al.* reported high substitutional carbon incorporation efficiencies at growth temperatures up to 600 °C using methylsilane ($SiCH_6$), where it was speculated that C may be incorporated into the lattice without breaking the preformed Si–C bond.

In summary, we have grown pseudomorphic $Si_{1-y}C_y$ layers with $y = 0.008$ by evaporation of Si in a graphite crucible at an effusion cell temperature of 1400 °C. Evaporation from the single silicon–graphite source results in an energy barrier to the formation of interstitial Si–C defects of $\Delta = 1.0 \pm 0.2$ eV compared to the separate evaporation of graphite and silicon with $\Delta = 0.50$ eV.[5] An infrared absorption mode observed near 725 cm^{-1} was attributed to a transitional Si–C phase intermediate between that of C diffusing from substitutional lattice sites to the formation of SiC precipitates.

The authors gratefully acknowledge support from ONR Grant No. N00014-93-1-0393 and DARPA Contract No. F49620-96-C-0006. They thank P. Thompson and M. Fatemi of the Naval Research Laboratory for the x-ray rocking curve measurements.

[1] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, Appl. Phys. Lett. **60**, 3035 (1992).
[2] H. J. Osten, E. Bugiel, and P. Zaumseil, Appl. Phys. Lett. **64**, 3440 (1994).
[3] K. Eberl, K. Brunner, and W. Winter, Thin Solid Films **294**, 98 (1997).
[4] J. Tersoff, Phys. Rev. Lett. **74**, 5080 (1995).
[5] H. J. Osten, M. Kim, K. Pressel, and P. Zaumseil, J. Appl. Phys. **80**, 6711 (1996).
[6] J. Drowart and G. De Maria, in *Silicon Carbide*, edited by J. R. O'Connor and J. Smiltens (Pergamon, London, 1960), p. 16.
[7] J. Drowart, G. De Maria, and M. Inghram, J. Chem. Phys. **29**, 1015 (1958).
[8] B. A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R. G. Wilson, P. R. Berger, and J. Kolodzey, J. Electron. Mater. **25**, 297 (1996).
[9] P. D. Zavitsanos and G. A. Carlson, J. Chem. Phys. **59**, 2966 (1972).
[10] J. W. Strane, H. J. Stein, S. R. Lee, B. L. Doyle, S. T. Picraux, and J. W. Mayer, Appl. Phys. Lett. **63**, 2786 (1993).
[11] J. W. Strane, H. J. Stein, S. R. Lee, S. T. Picraux, J. K. Watanabe, and J. W. Mayer, J. Appl. Phys. **76**, 3656 (1994).
[12] T. Kimura, S. Shigeru, and S. Yugo, Thin Solid Films **81**, 319 (1981).
[13] J. Mi, P. Warren, P. Letourneau, M. Judelewicz, M. Gailhanou, M. Dutoit, C. Dubois, and J. C. Dupuy, Appl. Phys. Lett. **67**, 259 (1995).

1.3 μm photoresponsivity in Si-based $Ge_{1-x}C_x$ photodiodes

Xiaoping Shao, S. L. Rommel, B. A. Orner, H. Feng, M. W. Dashiell, R. T. Troeger, J. Kolodzey, and Paul R. Berger[a]
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716-3130

Thomas Laursen
Center for Solid State Science, Arizona State University, Tempe, Arizona 85287-1704

(Received 2 December 1997; accepted for publication 12 February 1998)

$Ge_{1-x}C_x$/Si heterostructure photodiodes with nominal carbon percentages ($0 \leqslant x \leqslant 0.02$), which exceed the solubility limit, were grown by solid source molecular beam epitaxy on n-type (100) Si substrates. The p-$Ge_{1-x}C_x$/n-Si photodiodes were fabricated and tested. The p-$Ge_{1-x}C_x$/n-Si junction exhibits diode rectification with a reverse saturation current of about 10 pA/μm^2 at -1 V and high reverse breakdown voltage, up to -80 V. A significant reduction in diode reverse leakage current was observed by adding C to Ge, but these effects saturated with more C. Photoresponsivity was observed from these Si-based p-$Ge_{1-x}C_x$/n-Si photodiodes at a wavelength of $\geqslant 1.3$ μm, compatible with fiber optic wavelengths. External quantum efficiency of these thin surface-normal photodetectors was measured up to 2.2%, which decreased as the carbon percentage was increased. © *1998 American Institute of Physics.* [S0003-6951(98)00915-2]

Integration of cost effective photonic devices with mature Si-based microelectronic technology remains an area of intense research.[1–3] Current optoelectronic communication systems are based upon expensive but high performance III-V compound optoelectronic components and often cumbersome hybrid integration schemes. One such effort to realize affordable and easily mass-produced optoelectronic integrated circuits (OEICs) which could potentially impact the consumer market involves Si-based optoelectronic devices. These devices need to be monolithically integrated with Si-based microelectronics and operate in the infrared wavelength region, suitable for fiber-optic communications.

One such candidate is Group IV semiconductor alloys and heterostructures.[1–3] In the past decade, most work on Group IV semiconductors has focused on strained Si-rich SiGe alloys on Si substrates. To achieve an efficient photoresponse at the 1.3 μm wavelength region with SiGe/Si photodetectors, the Ge concentration should be in excess of 35%. However, the 4% lattice mismatch between Si and Ge constrains the design of SiGe heterojunction devices because of the limited critical thickness for pseudomorphic SiGe epilayers on a Si platform.[2]

Incorporation of C into Si and Ge to form $Ge_{1-x}C_x$ and the ternary alloy $Si_{1-x-y}Ge_xC_y$ makes it possible to reduce the compressive strain between the SiGe epilayer and the Si substrate, and therefore to control the strain about the lattice matching condition for Si substrates while allowing adjustable band gaps.[4–6] However, due to the low solubility of C in Ge (10^8 atoms/cm^3 at the melting point of Ge),[7] it is considerably difficult to obtain large fractions of C in Ge. $Ge_{1-x}C_x$ is predicted to be thermodynamically unstable in solid state form,[8] decomposing into its segregated components under zero pressure. But, nonequilibrium growth techniques, such as molecular beam epitaxy (MBE) or rapid thermal chemical vapor deposition (RTCVD), can synthesize $Ge_{1-x}C_x$ alloys with small percentages of carbon (under 3% by MBE[9] and 5% by RTCVD).[10]

This work has concentrated on the binary $Ge_{1-x}C_x$ alloy. Previous research on the $Ge_{1-x}C_x$ system has centered mostly upon its material properties. Our own investigations explored processing issues[11] and now device performance, using these new Group IV alloys. In this letter, we report on the fabrication and characterization of p-$Ge_{1-x}C_x$/n-Si heterojunction photodiodes grown by MBE with ($0 \leqslant x \leqslant 0.02$), nominally. Our diodes demonstrate that rectification and diode leakage significantly improves by adding C to Ge. Photoresponsivity to 1.3 μm light, suitable for fiber optic systems, was observed for the p-$Ge_{1-x}C_x$/n-Si photodiodes which were based upon a Si platform.

The normal-incidence heterojunction p-n photodiodes were fabricated from p-$Ge_{1-x}C_x$ epilayers grown on (100) n-Si substrates by MBE in an EPI 620 system. Details of the $Ge_{1-x}C_x$ growth are described elsewhere.[9] The carrier concentration of the n-Si substrates was 10^{14}–10^{15} cm^{-3}. The $Ge_{1-x}C_x$ bulk epilayers were *in situ* doped p type by a concurrent B flux, using an effusion cell loaded with pure B in a pyrolytic graphite crucible and all epilayers were about 0.6 μm thick. Information on the composition, doping and thickness of the p-$Ge_{1-x}C_x$ epilayers used in this study is listed in Table I. The nominal C compositions were determined from the growth conditions, which were calibrated from other samples.

The substrate temperature during MBE growth was kept constant at 400 °C. This temperature produces layer-by-layer growth of $Ge_{1-x}C_x$ to avoid growth front roughness, which occurs at elevated temperatures.[12] Also, this growth temperature minimizes outdiffusion of the B dopant into the Si substrate, and junction displacement. These Ge-rich layers were confirmed to be single crystal with good crystallinity away from the interface by channeling Rutherford backscattering (RBS). For epilayers greatly exceeding the critical thickness, it is expected that the epilayers should be relaxed with a high

---

[a] Electronic mail: pberger@ee.udel.edu

TABLE I. Selected information on the material properties, including composition, doping level, and epilayer thickness of the samples used in this study along with representative diode characteristics of the p-$Ge_{1-x}C_x$/n-Si heterojunction photodiodes, including dark current at selected bias points, ideality factor and external quantum efficiency (EQE) at 1.3 μm.

Sample Number	SGC174	SGC173	SGC175	SGC236
Nominal composition	$Ge_{0.992}C_{0.008}$	$Ge_{0.986}C_{0.014}$	$Ge_{0.98}C_{0.02}$	Ge
P-type doping level (cm^{-3})	2.7×10^{18}	2.1×10^{18}	1.5×10^{18}	2.0×10^{18}
Thickness (nm)	590 (RBS)[a]	610 (RBS)[a]	645 (RBS)[a]	600 (GC)[b]
Dark current (pA/μm^2)				
@−1 V	14	12	10	414
@−20 V	73	34	46	1800
Ideality factor	1.08	1.07	1.05	1.19
External quantum efficiency (%)				
@1.3 μm	2.2	2.0	1.3	2.2
corrected	2.8	2.5	1.6	2.8

[a] RBS: determined by RBS measurements.

[b] GC: estimated from growth condition.

density of defects, including misfit and threading dislocations. This was confirmed by transmission electron microscopy (TEM) in cross-sectional views on similar layers, which showed dislocation densities about 10^{10} cm^{-2}.

$Ge_{1-x}C_x$ mesas, which were 97 μm in diameter, were formed by photolithography and wet etching into the Si substrate, using a $H_3PO_4:H_2O_2:H_2O$ etchant.[13] A second photolithography defined the annular ohmic contact (94 μm outer diameter and 54 μm inner diameter), using standard liftoff technology. A Ti/Au bilayer was deposited by electron beam evaporation atop the mesa. The backside ohmic contact on the Si substrate was formed by e-beam evaporation of Ti/Au on the entire backside of the substrate. The samples then underwent heat treatments in a forming gas (15% H_2-N_2) ambient at 300 °C for 30 s in a Heatpulse 610 rapid thermal annealing (RTA) furnace. Adjacent to the diodes were transmission line method (TLM) test structures for measuring the ohmic contact resistance of the $Ge_{1-x}C_x$ contacts. The metal contacts were determined to be ohmic, with a specific contact resistance of under 2×10^{-6} Ω cm^2.

The current–voltage ($I-V$) characteristics of the p-$Ge_{1-x}C_x$/n-Si diodes were measured and are shown in Fig. 1. The $Ge_{1-x}C_x$/Si structures exhibit diode rectification. All the diodes exhibited reverse breakdown voltage of −60 to −80 V, irrespective of the C concentration varying from 0% to 2%. The measured reverse saturation currents of the three p-$Ge_{1-x}C_x$/n-Si diodes at −20 V and −1 V are in the range of 34–74 pA/μm^2 and 10–14 pA/μm^2, respectively (Table I). No clear dependence of leakage current on the C concentration was observed for any of the p-$Ge_{1-x}C_x$/n-Si diodes. This measured leakage current is comparable to a reported leakage current density of 70 pA/μm^2 for a $Si_{0.385}Ge_{0.600}C_{0.015}$ p-i-n photodetector.[14] The p-Ge/n-Si diode (SGC236) control diode with 0% C exhibits a much larger reverse saturation current, as shown in Fig. 1, than diodes which include C in the active region. This large change of reverse saturation currents between $Ge_{1-x}C_x$ diodes with C and without C agrees with earlier results.[15] The C may improve the Ge/Si interface quality and consequently the diode properties, but the small differences observed among the three p-$Ge_{1-x}C_x$/n-Si diodes could be due to the small variation in C composition. Therefore, a small amount of C can have a profound impact on leakage, but adding more C leads to diminishing returns.

The I-V curves of the p-$Ge_{1-x}C_x$/n-Si diodes (Fig. 1) all show a low turn on voltage of only 0.15–0.2 V for the p-n $Ge_{1-x}C_x$/Si heterojunctions regardless of C composition. A large series resistance in the low-doped Si substrate made extrapolation of the ideality factor, but an improvement of diode properties by adding C to Ge was observed. A summary of the diode I-V characteristics is listed in Table I.

Photoresponsivity was observed for all the $Ge_{1-x}C_x$/Si heterojunction diodes using laser excitation at a wavelength of 1.3 μm, which is below the Si band gap ($\lambda_g=1.1$ μm). The laser beam was focused within the inner circle of the annular contact onto the bare $Ge_{1-x}C_x$ surface. Photocurrents were measured in these diodes which lack any antireflection coatings and are shown in Fig. 2. The external quantum efficiencies (EQE) were calculated (see Table I). To account for the fraction of the incident laser spot falling outside the central window area and is reflected off the annular metal contact, a correction factor of about 0.8 was multiplied by the total incident light power, assuming a two-

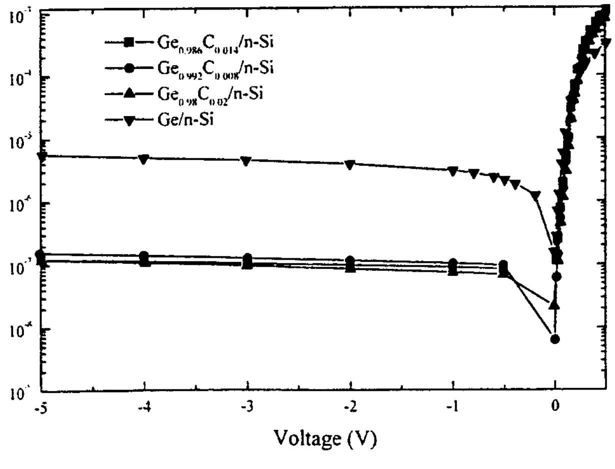


FIG. 1. The measured I-V characteristics (semi-log form) of the p-$Ge_{1-x}C_x$/n-Si diodes with 0%, 0.8%, 1.4%, and 2.0% carbon.

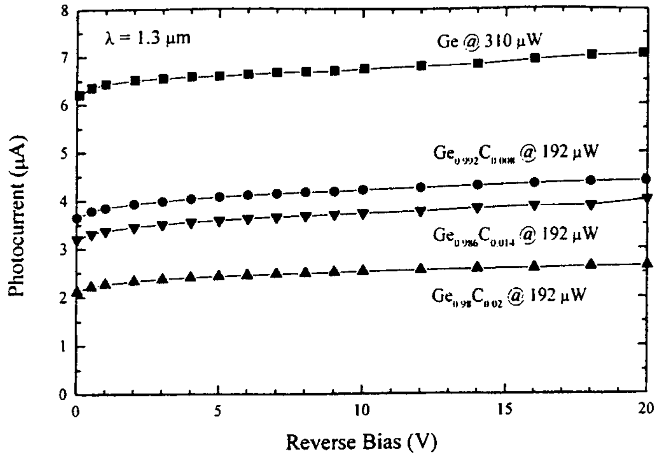


FIG. 2. The measured photocurrent for the p-$Ge_{1-x}C_x/n$-Si photodiodes with 0%, 0.8%, 1.4%, and 2% carbon and incident 1.3 μm illumination.

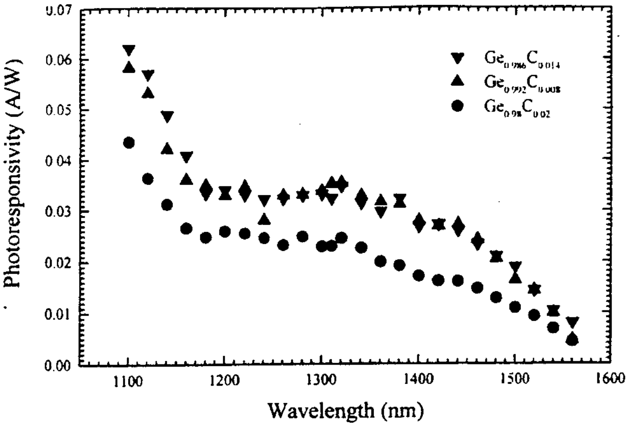


FIG. 3. The spectral photoresponsivity of the p-$Ge_{0.992}C_{0.008}/n$-Si, p-$Ge_{0.986}C_{0.014}/n$-Si, and p-$Ge_{0.98}C_{0.02}/n$-Si photodiodes. Note the elevated response at 1.1 μm due to the Si substrate.

dimensional (2D) Gaussian distribution for the incident light. The corrected EQEs are given in Table I.

Spectral dependence of the photoresponsivity was measured, as shown in Fig. 3. Note the elevated response for wavelengths below 1.1 μm where the absorption edge of the Si substrate becomes prominent. It is expected that there is a dependence of 1.3 μm photoresponsivity on C concentration, because substitutional C will increase the band gap of the $Ge_{1-x}C_x$ alloys and cause a blueshift of the absorption edge. This was shown with absorption studies for $Ge_{1-x}C_x$ (Ref. 16) and Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys.[17] From these studies the $Ge_{1-x}C_x$ band gap is expected to have a +63 meV/%C dependence. A significant drop in photoresponsivity for the p-$Ge_{0.98}C_{0.02}/n$-Si photodiodes compared to other p-$Ge_{1-x}C_x/n$-Si photodiodes was observed over the infrared region. Therefore, there may be a nonlinear dependence on EQE with increasing C. Also, a flat response is indicated between 1.2 and 1.3 μm which could be attributed to a defect level at 0.767 eV, a possible Si P-line, previously discovered in photoluminescence measurements of similar $Ge_{1-x}C_x$/Si layers.[13] However, it is clear that a small amount of C acts to significantly improve leakage currents without significantly affecting the 1.3 μm photoresponsivity.

The results of 1.3%–2.2 % quantum efficiency are significant not only because they are greater than the best result reported in normal incidence SiGeC p-i-n photodetectors, but also because the diodes did not have an anti-reflection (AR) coating and the light collection region is confined to a narrow region at the p^+-$Ge_{1-x}C_x/n^-$-Si one-sided junction. This leads to a much shorter collection length within the p^+-$Ge_{1-x}C_x$ depletion region and within a diffusion length than previous SiGeC p-i-n photodetectors which reported an EQE of ~1% at 1.3 μm, using an 800 Å thick active region.[14]

In conclusion, p-n photodiodes made from a heterojunction of epitaxial p-type $Ge_{1-x}C_x$ with nominal carbon percentages ($0 \leq x \leq 0.02$) on an n-type Si substrate were demonstrated. The I-V curves indicate that the p-Ge/n-Si diode exhibits a much larger reverse saturation current (~400 pA/μm^2 at −1 V) than that of a comparable p-$Ge_{1-x}C_x/n$-Si diode (10 pA/μm^2 at −1 V), suggesting that C could improve the $Ge_{1-x}C_x$ epilayer quality and suppress photodiode dark currents. High reverse breakdown voltages in excess of −60 and some up to −80 V are observed in all of the p-$Ge_{1-x}C_x/n$-Si diodes, including the p-Ge/n-Si diode. Despite the large number of dislocations and defects at the heterojunction, photoresponses with a measured EQE of 1.3%–2.2% from the $Ge_{1-x}C_x$ epilayers were observed by 1.3 μm laser excitation, and allow fiber-optic compatible photonics on a Si platform. There may be some nonlinear dependence of EQE on carbon concentration.

The authors wish to thank J. O. Olowolafe for useful discussions. This work was supported by the National Science Foundation (Grant No. ECS-9624160) and by DARPA (sponsored Research Agreement with Texas Instruments, Inc. under Grant No. SRA-3312665).

[1] R. A. Soref, Proc. IEEE **81**, 1687 (1993).
[2] R. People, IEEE J. Quantum Electron. **QE-22**, 1696 (1986).
[3] J. C. Bean, Proc. IEEE **80**, 571 (1992).
[4] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, Appl. Phys. Lett. **60**, 3033 (1992).
[5] B. Dietrich, H. J. Osten, H. Rücker, M. Methfessel, and P. Zaumseil, Phys. Rev. B **49**, 17 185 (1994).
[6] F. J. Guarin, S. S. Iyer, A. R. Powell, and B. A. Ek, Appl. Phys. Lett. **68**, 3608 (1996).
[7] R. I. Scace and G. A. Slack, J. Chem. Phys. **30**, 1551 (1959).
[8] O. F. Sankey, A. A. Demkov, W. T. Petuskey, and P. F. McMillan, Modelling Simul. Mater. Eng. **1**, 741 (1993).
[9] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, Appl. Phys. Lett. **67**, 1865 (1995).
[10] M. Todd, J. Kouvetakis, and D. Smith, Appl. Phys. Lett. **68**, 2407 (1996).
[11] X. Shao, S. L. Rommel, B. A. Orner, P. R. Berger, and J. Kolodzey, IEEE Electron Device Lett. **18**, 7 (1997).
[12] X. Shao, Ph.D. thesis, University of Delaware, 1997.
[13] A. S. T. Khan, Ph.D. thesis, University of Delaware, 1996.
[14] F. Y. Huang and K. L. Wang, Appl. Phys. Lett. **69**, 2330 (1996).
[15] A. Fukami, K. Shoji, T. Nagano, and C. Y. Yang, Appl. Phys. Lett. **57**, 2345 (1990).
[16] B. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R. G. Wilson, P. R. Berger, and J. Kolodzey, J. Electron. Mater. **25**, 297 (1996).
[17] B. A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J. W. Mayer, and J. Spear, Appl. Phys. Lett. **69**, 2557 (1996).

APPLIED PHYSICS LETTERS VOLUME 72, NUMBER 16 20 APRIL 1998

The effect of composition on the thermal stability of $Si_{1-x-y}Ge_xC_y/Si$ heterostructures

L. V. Kulik,[a] D. A. Hits, M. W. Dashiell, and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

(Received 14 November 1997; accepted for publication 17 February 1998)

The thermal stability of molecular beam epitaxy grown $Si_{1-x-y}Ge_xC_y/Si$ heterostructures ($0 \leqslant x < 0.30$, $y \sim 0.008$) was studied using infrared absorption spectroscopy. The local vibrational mode of C in Si and $Si_{1-x-y}Ge_x$ was used to quantify the loss of C atoms from substitutional sites with high temperature annealing. The activation energy ($E_a = 4.9$ eV) for the loss of substitutional C achieved a maximum for the strain compensated alloy ($x \sim 0.1$). An additional increase of Ge content resulted in a rapid decrease in E_a, which was found to be 3.4 eV for $x \sim 0.27$. The nonmonotonic behavior of E_a on Ge content is explained by the effect of the interface strain between the epitaxial layer and Si substrate. © *1998 American Institute of Physics.* [S0003-6951(98)00716-5]

Band-gap tailoring and lattice matching of $Si_{1-x-y}Ge_xC_y/Si$ heterostructures have the potential to improve the performance and capabilities of Si based optoelectronics.[1] Although remarkable progress in molecular beam epitaxy (MBE),[1,2] chemical vapor deposition (CVD),[3] and solid phase epitaxy (SPE)[4] of $Si_{1-x-y}Ge_xC_y/Si$ heterostructures has been achieved, important questions concerning growth kinetics and thermal stability are not yet fully understood. These growth techniques occur far from equilibrium forming metastable alloys with C concentrations exceeding the equilibrium solubility by up to four orders of magnitude. The magnitude of the C supersaturation in these alloys imposes limitations on thermal processing, beyond which the alloy will relax to its equilibrium state. Under high temperature treatment, the relaxation of metastable $Si_{1-y}C_y$ alloys occurs by the formation of thermodynamically stable β-SiC precipitates of small grain size (3–5 nm).[5] However, postgrowth device processing of these heterostructures requires that the structural properties and alloy compositions remain stable during heat treatments. There have been a number of reports on the measurements of thermal stability of $Si_{1-y}C_y/Si$ heterostructures,[5–7] but little is known of $Si_{1-x-y}Ge_xC_y/Si$ alloys.[7] We report on the effect of Ge on the thermal stability of $Si_{1-x-y}Ge_xC_y/Si$ heterostructures in the most technologically interesting region of Ge contents ($0 \leqslant x < 0.30$).

$Si_{1-x-y}Ge_xC_y$ layers were grown by MBE on *p*-type float zone Si(100) substrates. The Si–C beam was formed by thermally heating a float zone refined silicon ingot contained in a pyrolytic graphite crucible. Zone refined intrinsic Ge was evaporated from a pyrolytic BN crucible. The substrate temperature during growth was 400 °C. All layers were grown approximately 150 nm thick and were recrystallized at 650 °C for 30 min prior to the annealing experiment and compositional analysis. The recrystallization temperature and time were chosen so to maximize the integrated absorption of the local vibrational mode (LVM) of substitutional C in the layer. The procedure is similar to that used by Strane during the SPE growth of implanted $Si_{1-x-y}Ge_xC_y$ layers which resulted in nearly all C atoms occupying substitutional sites.[4] X-ray diffraction and Raman spectroscopy were applied to quantify the amount and uniformity of substitutional C and Ge in our samples. X-ray rocking curve analysis of the symmetric (004) and asymmetric (224) reflections revealed that the $Si_{1-x-y}Ge_xC_y$ layers with $x<0.2$ were fully strained. Further details on the MBE growth conditions can be found in Ref. 8

Samples were cut into 1 cm^2 pieces and annealed in a quartz furnace purged by pure N_2 at atmospheric pressure in the temperature region from 750 to 1000 °C for various times. Fourier transform infrared spectroscopy (FTIR) was used to analyze the amount of substitutional C during annealing through the integrated absorbance of the local vibrational mode of C in Si.

The left inset in Fig. 1 displays the room temperature FTIR spectra of the $Si_{0.992}C_{0.008}$ (solid line). The full width at half maximum (FWHM) of the substitutional C local vibration mode (C-LVM) at 607 cm^{-1} of $Si_{0.992}C_{0.008}$ is about 14 cm^{-1}, close to that found by Strane *et al.*[5] The dashed line shows the absorbance of the $Si_{0.722}Ge_{0.27}C_{0.008}$ alloy, illustrating the spectral shift of the C-LVM to lower energies due to the effect of Ge on the bonding strength.

Figure 1 shows absorption spectra of $Si_{0.892}Ge_{0.10}C_{0.008}$ annealed at 908 °C for different times. The spectra consists of two main lines at 602 and at 815 cm^{-1}. The asymmetric line at 815 cm^{-1} with FWHM ~66 cm^{-1} is associated with the vibration of coherent SiC precipitates.[5] The line at 602 cm^{-1} is the vibrational mode of substitutional C. With progressive annealing, the absorbance at 602 cm^{-1} decreases while that of the 815 cm^{-1} mode increases showing that C atoms transfer from substitutional sites to SiC precipitates. Samples were annealed until the C vibrational mode disappeared (the uppermost spectrum in Fig. 1). The spectrum without the C mode was then subtracted from the others as a reference to remove all irrelevant spectral features when determining the integrated absorbance of the C-LVM. This procedure was especially important for $Si_{1-x-y}Ge_xC_y/Si$ het-

---

[a] On leave from The Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow Region, Russia.

© 1998 American Institute of Physics

Appendix 35

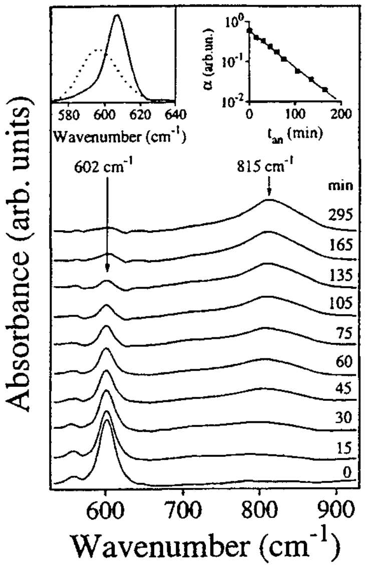


FIG. 1. Effect of annealing at 908 °C on FTIR spectra of $Si_{0.892}Ge_{0.10}C_{0.008}$/Si heterostructure. Annealing times (t_{an}) are shown on the right side of the figure. The right inset shows the integrated absorption coefficient of substitutional C vibration mode (α) vs t_{an}. Solid line is the best fit of the experimental points. The left inset shows the FTIR spectra of substitutional C vibration mode in $Si_{1-x-y}Ge_xC_y$/Si heterostructures with $x=0$ (solid line) and 0.27 (dashed line).

erostructures because of the strong difference between refractive indexes of the layers and the Si substrates.

The integrated absorption of the C-LVM (α) for $Si_{0.892}Ge_{0.10}C_{0.008}$ alloys versus annealing time at 908 °C is presented in the top right inset to Fig. 1. It can be fit with an exponential decay function

$$\alpha=\alpha_0\times\exp\left\{-\frac{t}{\tau}\right\}, \tag{1}$$

with α_0 and τ being the initial absorption coefficient and decay constant, respectively. To ensure the accuracy of the fitting procedure, α was measured over a duration in which its magnitude decreased by at least one order of magnitude; see the right inset in Fig. 1. Similar to Strane *et al.*,[5] we found a rapid nonexponential decrease in α during the initial annealing steps at annealing temperature less than 900 °C, which was followed by a continuous exponential decay of α for the remainder of the annealing sequence. X-ray diffraction measurements, however, showed the layer strain decreased exponentially for the entire annealing sequence with the same decay constant as the exponential part of α decay. We believe that the initial drop of α is the result of the formation of a transitional Si–C phase partially preserving the layer strain, which precedes the precipitation of β-SiC.[5,8] In fact, during the initial drop of α, no characteristic β-SiC absorption was observed in FTIR spectra. Therefore, only the onset of the exponential part of the α decay curve will be considered further.

Figure 2 presents the Arrhenius plot relating the measured decay constant for loss of substitutional C versus annealing temperature for $Si_{1-x-y}Ge_xC_y$/Si heterostructures with $x=0$, 0.1, and 0.23. For each value of Ge content, $1/\tau$ reveals thermally activated behavior over the entire temperature region

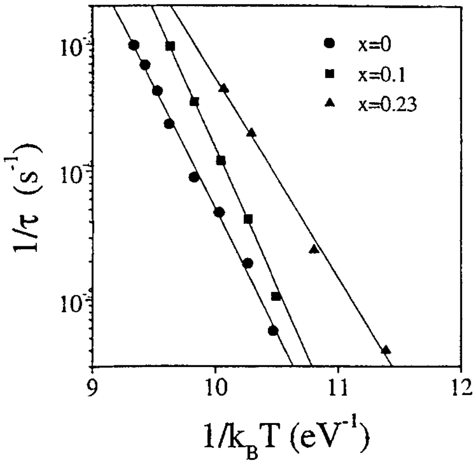


FIG. 2. Arrhenius plot of the rate of substitutional C loss for three different Ge contents ($x=0$, 0.1, 0.23). Solid lines are the best fits of the experimental points.

$$\frac{1}{\tau}=r_0\times\exp\left\{-\frac{E_a}{k_BT}\right\}. \tag{2}$$

The magnitude of E_a for the $Si_{0.992}C_{0.008}$ alloy found from the best fit of the experimental points is 4.45±0.15. The E_a for the $Si_{0.892}Ge_{0.1}C_{0.008}$ layer is 4.9±0.2 eV. The values of E_a for $Si_{1-y}C_y$/Si and $Si_{1-x-y}Ge_xC_y$/Si ($x\sim0.1$) heterostructures are in good agreement with the prior results measured in the high temperature annealing region (1000–1130 °C) for this same composition (4.5 and 5.3 eV, respectively),[7] and far exceeds the activation energy of C diffusion in bulk Si near the C solubility limit (3.1 eV).[9] A discrepancy between our activation energy for $Si_{1-y}C_y$ versus a lower value obtained by Fischer *et al.* ($E_a=3.3$ eV) may be due to surface oxidation during their annealing sequence.[6] Due to volume considerations, the formation of SiO_2 will inject one silicon self-interstitial for each of the two-oxygen atoms.[10] Carbon diffusion is enhanced in the presence of a supersaturation of Si self-interstitials, as C diffusion is believed to be dominated by a mobile interstitial C–Si pair.[10] For the temperature range of 800–1100 °C, Ladd *et al.* measured an increase in C diffusivity by approximately one to two orders of magnitude during annealing in an oxidizing ambient and during a phosphorus in-diffusion when compared to annealing in a nitrogen ambient.[11] The former two environments produce a Si self-interstitial concentration in excess of the equilibrium concentration. The surface oxide observed during Fischer's work may have been a source of excess Si self-interstitial defects, resulting in increased C diffusivity and thus a lower barrier to loss of substitutional C. During our experiments, we took great care to ensure all annealing was performed in a pure N_2 environment. Combining our results with those of Ref. 7, a unitary activation behavior for the loss of substitutional C for $Si_{1-y}C_y$/Si and $Si_{1-x-y}Ge_xC_y$/Si ($x\sim0.1$) heterostructures in the entire temperature region 830–1130 °C is concluded.

As shown in Fig. 3, the E_a vs Ge content achieves maximum at $x\sim0.1$, and decreases rapidly at higher x. Such a behavior can be understood as the competition of opposing contributions to E_a: (i) the mismatch between Si–Si and Si–C bonds becomes larger with Ge content which acts to decrease E_a,[12,13] and (ii) the interface strain between the

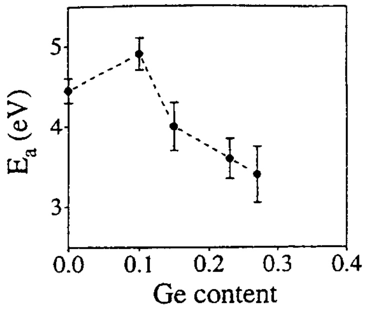


FIG. 3. Activation energy of substitutional C loss vs Ge content for annealed $Si_{1-x}Ge_xC_{0.008}$. The maximum at $x=0.1$ occurs for the strain compensated heterostructure.

layer and the Si substrate decreases as the small addition of Ge to $Si_{1-y}C_y$ compensates the tensile strain in the layer[2] thus increasing E_a. In our case the strain compensation is attained at $x\sim 0.1$, which was verified by x-ray diffraction measurements. At $x\sim 0.1$ when the C atoms leave substitutional lattice sites, the strain energy increases by a value ΔE, as opposed to in $Si_{1-y}C_y$ layers where the strain energy decreases with loss of substitutional C. From continuous elastic theory the contribution from interface strain at $x\sim 0.1$ $(2\Delta E)$ could be estimated as ~ 0.1 eV per C atom.[14] This value is in reasonable agreement with the experimental results if one takes into account the limitations of the elastic theory approach for SiGeC alloys with large mismatch between Si–Si(Ge), and Si(Ge)–C bonds.[2,13]

In conclusion, we present temperature stability measurements of $Si_{1-x-y}Ge_xC_y$/Si and $Si_{1-y}C_y$/Si MBE grown heterostructures with $(0\leqslant x<0.30,\ y\sim 0.008)$. The ternary alloy $Si_{1-x-y}Ge_xC_y$/Si heterostructure with small $x\sim 0.1$ was found to have higher activation energy for substitutional C loss than for the binary $Si_{1-x}C_y$ alloy. The activation energy achieves a maximum in the structure with zero interface strain between the Si substrate and $Si_{1-x-y}Ge_xC_y$ layers $(x\sim 0.10$ and $y=0.008)$ and rapidly decreases as the Ge fraction is further increased. Thus, ternary unstrained layers grown on Si substrate might hold the greatest potential for device processing at elevated temperatures for Si based heterostructure technology.

The authors acknowledge support from ONR Grant No. N00014-93-1-0393, DARPA Contract No. F49620-96-C-0006, and ARO Grant No. DAAH04-95-1-0625. They also thank P. Thompson and M. Fatemi of the Naval Research Laboratory for x-ray rocking curve measurements, and G. Watson of the Department of Physics and Astronomy of the University of Delaware for Raman measurements. Special thanks to H. J. Osten of the Institute of Semiconductor Physics in Frankfurt (Oder) for useful discussions.

[1] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, Appl. Phys. Lett. **60**, 3035 (1992).
[2] B. Dietrich, H. J. Osten, H. Rücker, M. Methfessel, and P. Zaumseil, Phys. Rev. B **49**, 17185 (1994).
[3] A. St. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewolt, Appl. Phys. Lett. **67**, 3915 (1995).
[4] J. W. Strane, H. J. Stein, S. R. Lee, B. L. Doyle, S. T. Picraux, and J. M. Mayer, Appl. Phys. Lett. **63**, 2786 (1993).
[5] J. W. Strane, H. J. Stein, S. R. Lee, S. T. Picraux, J. K. Watanabe, and J. M. Mayer, J. Appl. Phys. **76**, 3656 (1994).
[6] G. G. Fisher, P. Zaumseil, E. Bugiel, and H. J. Osten, J. Appl. Phys. **77**, 1934 (1995).
[7] P. Warren, J. Mi, F. Overney, and M. Dutoit, J. Cryst. Growth **157**, 414 (1995).
[8] M. W. Dashiell, L. V. Kulik, D. A. Hits, G. Watson, and J. Kolodzey, Appl. Phys. Lett. **72**, 833 (1998).
[9] R. C. Newman and J. Wakefield, in *Metallurgy of Semiconductor Materials*, edited by J. B. Schroeder (Interscience, New York, 1961), pp. 15 and 201.
[10] U. Gosele, Mater. Res. Soc. Symp. Proc. **59**, 419 (1986).
[11] L. A. Ladd and J. P. Kalejs, Mater. Res. Soc. Symp. Proc. **59**, 445 (1986).
[12] O. F. Sankey, A. A. Demkov, W. T. Petuskey, and P. F. McMillan, Modelling Simul. Mater. Sci. Eng. **1**, 741 (1993).
[13] P. C. Kelires, Phys. Rev. Lett. **75**, 1114 (1995).
[14] E. Kasper and H. J. Herzog, Thin Solid Films **44**, 357 (1977).

Appendix 36

ELSEVIER

Thin Solid Films 313–314 (1998) 172–176

Optical properties and band structure of $Ge_{1-y}C_y$ and Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys

K.E. Junge[a], N.R. Voss[a], R. Lange[a], J.M. Dolan[a], Stefan Zollner[a,*], M. Dashiell[b], D.A. Hits[b], B.A. Orner[b], R. Jonczyk[b], J. Kolodzey[b]

[a]*Ames Laboratory and Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA*
[b]*Department of Electrical and Computer Engineering, 140 Evans Hall, University of Delaware, Newark, DE 19716, USA*

Abstract

We measured the dielectric function of $Ge_{1-y}C_y$ and Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys from 1.6 to 5.2 eV using spectroscopic ellipsometry. These alloys were grown by molecular beam epitaxy at 600°C on (001) Si substrates. Analytic lineshapes fitted to numerically calculated derivatives of their dielectric functions determined the critical-point parameters of the E_1, $E_1+\Delta_1$, E_0', and E_2 transitions. The critical-point energies of the $Ge_{1-y}C_y$ alloys were found to be indistinguishable from those of bulk Ge. This indicates that the presence of C in these alloys has no detectable influence on the band structure. The amplitude of the ellipsometric spectra is much lower than for bulk Ge, which can be attributed to surface roughness and explained within the framework of the Kirchhoff theory of diffraction or using effective medium theory. The degree of surface roughness indicated by optical measurements was verified by atomic force microscopy. © 1998 Elsevier Science S.A.

Keywords: Optical properties; Dielectric function; Germanium; Carbon

---

1. Introduction

Group-IV alloys promise improved performance of Si-based devices without using smaller VLSI length scales [1]. A detailed understanding of their optical properties is needed for monitoring of epitaxial growth, but our knowledge is sketchy, confined to bulk $Si_{1-x}Ge_x$ data [2] and limited data for epitaxial films [3–5]. The introduction of C into $Si_{1-x}Ge_x$ reduces the strain (and associated defects) in pseudomorphic layers and provides an additional parameter for band structure engineering [6].

The E_1 and $E_1+\Delta_1$ critical points (CPs) of pseudomorphically strained $Si_{1-x}Ge_x$ films on Si can be understood using bulk data, if the biaxial strain is taken into account using deformation potentials [3–5]. Surprisingly (in contrast to theoretical predictions and photoluminescence measurements of the indirect gap), the same is also true for $Si_{1-y}C_y$ [7–14] and $Si_{1-x-y}Ge_xC_y$ alloys ($x \leq 0.3$, $y \leq 0.02$) [15]. The CP parameters for the ternary $Si_{1-x-y}Ge_xC_y$ alloys are difficult to interpret, since the C and Ge fractions are not known with sufficient accuracy (0.1%). A broadening of the CPs is observed with increasing disorder, particularly when adding C [15].

In this work, we use spectroscopic ellipsometry and atomic force microscopy (AFM) to study $Ge_{1-y}C_y$ and Ge-rich $Si_{1-x-y}Ge_xC_y$ on Si. The large lattice mismatch is relaxed by defects at the interface (e.g. dislocations). We have shown previously that the dielectric response of Ge on Si is similar to that of bulk Ge, since most misfit dislocations are localized near the interface [16–18]. We find that the CP energies of

---

[*] Corresponding author, Present address: SPS, Arizona Technology Laboratories, MD M360, 2200 West Broadway Road, Mesa, AZ 85202, USA. Tel.: +1 602 6554655; fax: +1 602 6555013; e-mail: R37595@email.sps.mot.com

0040-6090/98/$19.00 © 1998 Elsevier Science S.A. All rights reserved
PII S0040-6090(97)00806-7

Table 1
Composition and thickness (d) of samples studied. The ratio of substitutional to total C is given in parentheses, see Ref. [24]

Sample	Si fraction	Ge fraction	C fraction	d (μm)
SGC-32	0	0.97	0.03	0.582
SGC-31	0	0.98	0.02	0.139
SGC-30	0	0.99	0.01	0.136
SGC-70	0.11	0.88	0.01 ($\leq$ 0.25)	0.19
SGC-69	0.11	0.88	0.01 ($\geq$ 0.70)	0.14
SGC-68	0.11	0.88	0.01 ($\leq$ 0.25)	0.14
SGC-67	0.135	0.845	0.02 ($\leq$ 0.25)	0.14

$Ge_{1-y}C_y$ and Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys are not affected by C, in agreement with similar work by Krishnamurthy et al. [18]. Apparently, it is much more difficult to form Ge–C bonds than Si–C bonds [19]. These samples are very rough, a fact which can be described using the theory of Ohlídal and Lukeš [20] or effective medium theory.

2. Growth and characterization

The alloys were grown by molecular beam epitaxy (MBE) on Si (001) at 600°C ([21–24] and Hits et al., unpublished results). Ge was thermally evaporated in a pyrolytic boron nitride crucible at 1260°C, and Si in a pyrolytic graphite crucible at 1685°C. The C source consisted of a pyrolytic graphite filament. Growth rates were typically 0.06 μm/h.

Film compositions (Table 1) were measured by Rutherford backscattering (RBS). The amount of substitutional C in Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys was studied using ion channeling. In SGC-67, SGC-68, and SGC-70 less than 25%, but in SGC-69 at least 70% of the total C was found to be substitutional [24]. The probable C source for SGC-69 was the graphite crucible used for evaporating Si, because the C filament was turned off. The amount of substitutional C in $Ge_{1-y}C_y$ alloys is not known, but is expected to be low [18].

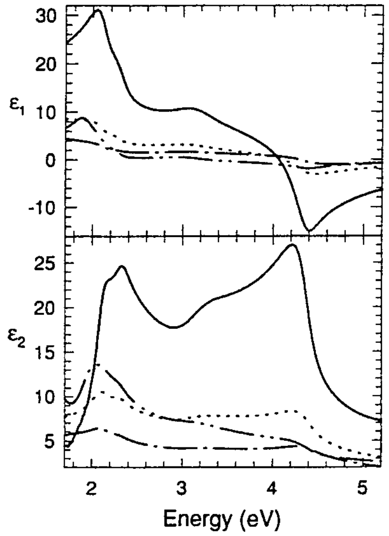


Fig. 1. Pseudodielectric function of $Ge_{1-y}C_y$ for $y=0.01$ (dotted), 0.02 (dot-dashed), 0.03 (double dot-dashed). Bulk Ge data (solid line) of Ref. [2] are shown for comparison.

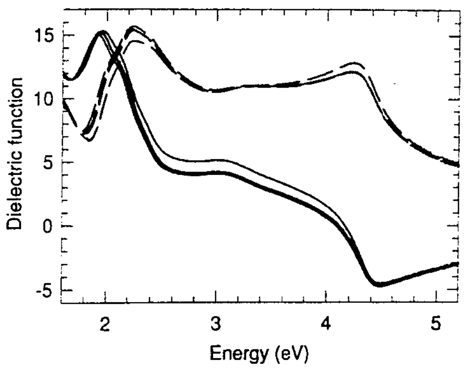


Fig. 2. Real (solid) and imaginary (dashed) part of $\langle\varepsilon\rangle$ for four $Si_{1-x-y}Ge_xC_y$ ($x\approx 0.9$, $y=0$–0.02) alloys.

3. Ellipsometry

The pseudodielectric functions [25] $\langle\varepsilon\rangle$ of these layers were measured ex situ from 1.6 to 5.2 eV with a spectroscopic ellipsometer of the rotating analyzer-type [26]. Fig. 1 shows $\langle\varepsilon\rangle$ for $Ge_{1-y}C_y$ in comparison with bulk Ge [2] (solid line). The $Ge_{1-y}C_y$ alloys resemble the spectrum of bulk Ge except for a lower amplitude attributed to surface roughness, which will be discussed later. $\langle\varepsilon\rangle$ of Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys is given in Fig. 2.

The spectra in Figs. 1 and 2 show a double-peak structure above 2 eV (the E_1 and the $E_1+\Delta_1$ doublet), a shoulder near 3 eV (E_0'), and a third peak near 4.2 eV (E_2). These peaks are due to CPs arising from direct band-to-band transitions at various regions in the Brillouin zone [27]. For a further analysis of these CPs, we numerically calculate the second derivative of $\langle\varepsilon\rangle$ with respect to photon energy (shown by symbols in Fig. 3) and perform a line shape analysis. We describe the CPs using a mixture of a 2D minimum and a saddle point [27]

$$\varepsilon(\omega)=C-A\ln(\hbar\omega-E_g-i\Gamma)\exp(i\phi), \qquad (1)$$

where $\hbar\omega$ is the photon energy, E_g the energy of the CP, Γ its broadening, A its amplitude (oscillator strength), and ϕ the phase angle describing the amount of mixing. The parameters obtained from the lineshape analysis for $Ge_{1-y}C_y$ and Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys are given in Table 2 and Table 3, respectively. For comparison, the parameters of bulk Ge from Viña et al. [27] and a commercial bulk Ge

Table 2
Critical point (CP) parameters of $Ge_{1-y}C_y$: amplitude (A), energy (E_g), broadening (Γ), and excitonic phase (Φ). The phase angle of the $E_1 + \Delta_1$ transition was forced to be the same as that of E_1

Sample	CP	A (1)	E_g (eV)	Γ (eV)	Φ (deg)
	E_1	1.27(7)	2.12(1)	0.071(3)	142(4)
SGC-32	$E_1 + \Delta_1$	1.1(1)	2.32(1)	0.090(4)	same
$y = 0.03$	E_0'	0.6(1)	2.98(2)	0.20(2)	−15(12)
	E_2	0.8(1)	4.37(1)	0.13(1)	181(9)
	E_1	0.57(6)	2.120(4)	0.066(3)	136(7)
SGC-31	$E_1 + \Delta_1$	0.4(1)	2.331(6)	0.09(1)	same
$y = 0.02$	E_0'	0.38(8)	3.00(3)	0.24(3)	−34(12)
	E_2	0.7(1)	4.37(1)	0.12(1)	120(8)
	E_1	1.3(2)	2.113(5)	0.064(4)	119(8)
SGC-30	$E_1 + \Delta_1$	1.2(3)	2.319(6)	0.09(1)	same
$y = 0.01$	E_0'	0.9(2)	3.03(3)	0.22(3)	−25(15)
	E_2	1.9(2)	4.36(1)	0.116(7)	145(6)
Bulk	E_1	5.5(3)	2.114(2)	0.057(2)	86(4)
$y = 0.00$	$E_1 + \Delta_1$	4.1(5)	2.314(3)	0.075(5)	same
(Eagle Picher)	E_0'	3.2(6)	3.05(2)	0.20(2)	−29(12)
Ref. [17]	E_2	8(1)	4.37(1)	0.104(9)	165(9)
Bulk (Viña et al. [27])					
Bulk	E_1		2.113(3)	0.06(1)	71(4)
$y = 0.00$	$E_1 + \Delta_1$		2.298(3)	0.07(2)	same
	E_0'		3.11		
	E_2		4.368(4)	0.109(9)	

⟨001⟩ sample (Eagle Picher) measured with our instrument are also given in Table 2.

The CP parameters for $Ge_{1-y}C_y$, see Table 2, are similar to those of bulk Ge with a few exceptions. Firstly, we note that the E_1 energy increases slightly with C concentration, but still remains within the error bars of the bulk energy. The C in these samples hardly affects the band structure. On the other hand, we would expect a 50 meV/%C shift for substitutional C [15]. Small amounts of substitutional C were also found by Krishnamurthy et al. [18]. Secondly, the CP broadenings increase with C incorporation due to the increasing disorder or random shear strains. The phase angle for the E_1 CP is large due to interference from the substrate. (We did not correct our spectra to take into account the finite film thickness, since this does not affect our conclusions.) In our analysis, we treat the spin-orbit splitting parameter Δ_1 as a free parameter (since it is a measure for the shear strain in the sample [15]) and find $\Delta_1 \approx 200$ meV, indicating little shear strain if any.

Table 3
Critical point (CP) parameters of $Si_{1-x-y}Ge_xC_y$

Sample	CP	A (1)	E_g (eV)	Γ (eV)	Φ (deg)
SGC-70	E_1	2.3(3)	2.180(7)	0.091(6)	73(9)
$x = 0.88$	$E_1 + \Delta 1$	0.8(2)	2.378(6)	0.08(2)	same
$y = 0.01$	E_0'	1.5(2)	2.98(2)	0.25(2)	−52(8)
	E_2	2.9(4)	4.36(1)	0.13(1)	138(9)
SGC-69	E_1	2.2(2)	2.183(7)	0.089(5)	78(9)
$x = 0.88$	$E_1 + \Delta_1$	0.8(3)	2.378(7)	0.09(2)	same
$y = 0.01$	E_0'	1.6(2)	2.98(3)	0.25(2)	−53(8)
	E_2	3.1(4)	4.36(1)	0.14(1)	139(7)
SGC-68	E_1	1.8(3)	2.20(1)	0.091(8)	87(11)
$x = 0.88$	$E_1 + \Delta_1$	0.9(4)	2.398(8)	0.09(2)	same
$y = 0.01$	E_0'	1.7(2)	2.99(2)	0.25(2)	−55(8)
	E_2	3.3(4)	4.36(1)	0.13(1)	135(8)
SGC-67	E_1	2.1(2)	2.210(7)	0.057(2)	81(11)
$x = 0.845$	$E_1 + \Delta_1$	0.9(3)	2.396(7)	0.09(2)	same
$y = 0.02$	E_0'	1.6(2)	2.97(2)	0.24(2)	−56(9)
	E_2	2.9(3)	4.36(1)	0.14(1)	140(7)

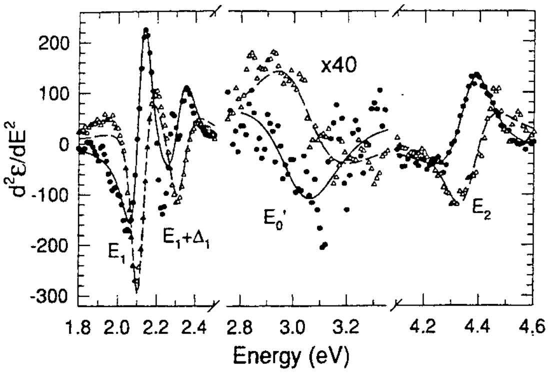


Fig. 3. Numerically calculated second derivatives of ε_1 (●) and ε_2 (△) for SGC-30 ($y = 0.01$). The lines give the best fit to Eq. (1) with the parameters in Table 2. The E_0' region (2.75–3.35 eV) was multiplied by 40.

The CP parameters for SGC-67, SGC-68, and SGC-70, see Table 3, can be explained by the binary $Si_{1-x}Ge_x$ system, because less than 25% of the total C is substitutional ([24], Hits et al., unpublished results). The E_1 energy, see Eq. 4b of Ref. [2], gives an effective Si concentration which is ~ 3% lower than given by the RBS data ([24], Hits et al., unpublished results). The broadening has increased and the amplitude has decreased, a fact which is attributed to surface roughness. The energy we measured for the broad E_0' CP is slightly lower than in the bulk within our uncertainty. The CP parameters of SGC-69 and SGC-70 are virtually identical. We conclude that the substitutional C in SGC-69 is not affecting E_1.

4. Surface roughness

The low amplitude of $\langle \varepsilon \rangle$ can be explained by surface roughness of the order of the optical penetration depth. The sum rule [28]

$$N_{eff} = \left[2m_0\varepsilon_0/(\pi e^2)\right] \int_{1.6\mathrm{eV}}^{5.2\mathrm{eV}} \omega\varepsilon_2(\omega)\mathrm{d}\omega \qquad (2)$$

can evaluate the effective electron density N_{eff} from

Table 4
Effective electron density N_{eff} and static dielectric constant $\varepsilon_1(0)$ calculated from $\varepsilon_2(\omega)$ using sum rules, see Eq. (2) and Ref. [28]

Sample	N_{eff} (e^-/atom)	$\varepsilon_1(0)$ (1)
Bulk	2.1	14
SGC-32	0.8	7
SGC-31	0.4	4
SGC-30	0.9	7

ε_2, where m_0 is the free electron mass. As given in Table 4, N_{eff} for $\text{Ge}_{1-y}\text{C}_y$ is less than half of the expected bulk Ge density. This can be understood by assuming that the epitaxial layer contains voids, which reduce the effective electron density. A similar sum rule can be derived for the static dielectric constant $\varepsilon_1(0)$ [28], see Table 4.

A quantitative description of the effect of surface roughness on $\langle \varepsilon \rangle$ was given by Ohlídal and Lukeš [20] (OL). This theory is based on the Kirchhoff theory of diffraction and gives the complex reflectance ratio for the rough surface as

$$\langle \rho \rangle = \frac{\rho_0 + \frac{1}{2}\tan^2\beta_0\left[r_{px}''(\theta_1) + r_{py}''(\theta_1)\right]/r_s(\theta_1)}{1 + \frac{1}{2}\tan^2\beta_0\left[r_{sx}''(\theta_1) + r_{sy}''(\theta_1)\right]/r_s(\theta_1)} \quad (3)$$

It depends on the reflectance ratio ρ_0 of a smooth surface, a roughness parameter ($\tan\beta_0$), and complex reflectances for s and p polarized light (r_j) as well as their derivatives with respect to the slope of the roughness. The roughness parameter is an aspect ratio related to the rms surface roughness (σ) and the correlation length (L) by $\tan\beta_0 = \sqrt{2}\,\sigma/L$ [20,29]. Results of this theory are only valid in the non-diffraction limit, where the wavelengths ($\lambda \ll 4\pi r_c \cos\theta$) must be shorter than the radius of curvature ($r_c \approx L^2/\sigma \approx 2$–$5$ μm), which implies $\lambda \ll 8$ μm in our case. The expected error in the results due to the Taylor series expansion is of the order of $(\sigma/L)^2$ because we only keep the quadratic term.

As shown in Fig. 4, the amplitude of $\langle \varepsilon \rangle$ decreases by increasing $\tan\beta_0$ in the OL theory. Similar results can be obtained within the effective medium approximation

$$\frac{\varepsilon_a - \langle \varepsilon \rangle}{\varepsilon_a + \kappa\langle \varepsilon \rangle} f_a + \frac{\varepsilon_b - \langle \varepsilon \rangle}{\varepsilon_b + \kappa\langle \varepsilon \rangle}(1 - f_a) = 0, \quad (4)$$

where κ is a constant of the order of unity related to the shape of the voids ($\kappa = 2$ for spherical voids), ε_a and ε_b the dielectric functions of the two components (Ge and voids), and f_a the Ge fraction in the film [29].

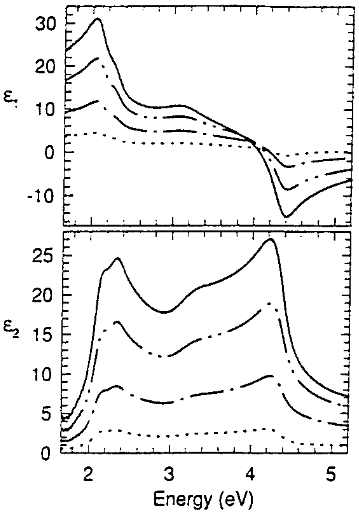


Fig. 4. Effects of surface roughness on the dielectric function of Ge (solid line, from Ref. [2]) calculated using Eq. (3) with the roughness parameters $\tan\beta_0 = 0.14$ (double dot-dashed), 0.24 (dot-dashed), 0.35 (dotted).

The OL theory considers the presence of an oxide layer by the identical film treatment (both boundaries are geometrically and statistically identical) [20]. The ellipsometric parameters characterizing the system are produced by the same procedure as for the simple rough surface. $\langle \rho \rangle$ is still given by Eq. (3), but ρ_0 is now the reflectance ratio for a smooth surface covered with an oxide with thickness t and the derivatives of the reflectances are more complicated. The dielectric function of SGC-30 is shown in Fig. 5 along with the best OL fit with and without an oxide layer. Including the oxide layer leads to a better fit and a smaller roughness parameter, but the oxide is a bit thicker than we would expect for a native oxide on a smooth surface, particularly for SGC-32. The fitted roughness ratio (σ/L) from ellipsometry, see Table 5, is shown along with measured AFM values. Considering our simple one-layer model, the agreement is reasonable.

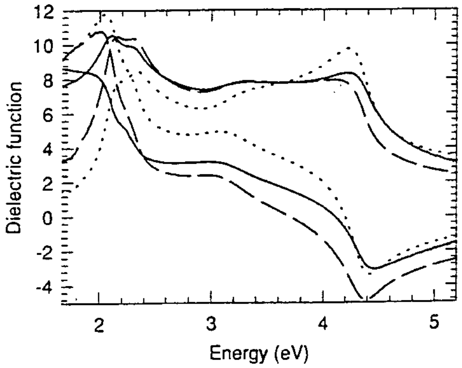


Fig. 5. Dielectric function of SGC-30 (solid line), the best fit without oxide (dotted) using Eq. (3) with $\tan\beta_0 = 0.24$, and with 34 Å of oxide and $\tan\beta_0 = 0.21$ (dashed).

Table 5
Rms surface roughness (σ) and correlation length (L) measured by atomic force microscopy. The measured ratio (σ/L) is compared to the fitted ellipsometry ratio with and without an oxide layer of thickness t, see Eq. (3)

Sample	AFM results			Ellipsometry data fit		
	σ (nm)	L (nm)	σ/L	With oxide		Without oxide
				σ/L	t (nm)	σ/L
SGC-70	21.2	269	0.08	0.11	22	0.14
SGC-69	19.3	258	0.07	0.11	21	0.13
SGC-68	21.2	269	0.08	0.11	14	0.13
SGC-67	20.5	269	0.08	0.11	21	0.14
SGC-31				0.21	39	0.22
SGC-30	34.9	263	0.13	0.15	35	0.17
SGC-32	9.7	225	0.04	0.13	96	0.21

5. Conclusions

The critical-point parameters of $Ge_{1-y}C_y$ and Ge-rich $Si_{1-x-y}Ge_xC_y$ alloys grown on Si (001) by MBE are similar to those of equivalent films not containing C. This indicates that C has no detectable influence on the energies of direct band-to-band transitions. The amplitude of $\langle \varepsilon \rangle$ of these alloys is lower than expected which can be explained by surface roughness or voids in the bulk (if present). The degree of surface roughness has been verified using atomic force microscopy.

Acknowledgements

Ames Laboratory is operated for US-DOE by ISU under Contract No. W-7405-Eng-82. The work at Ames was supported by the Director for Energy Research, Office of BES, and by the Iowa Space Grant Consortium. The work at Delaware was supported by AFOSR (F49620-95-0135), ARO (DAAH04-95-1-0625), DARPA (F49620-96-C-006), and ONR (N00014-93-1-0393). J.M.D. acknowledges support through a Bernice Black Durand Undergraduate Research Scholarship.

References

[1] B.S. Meyerson, Sci. Am. 270 (1994) 62.
[2] J. Humlíček, M. Garriga, M.I. Alonso, M. Cardona, J. Appl. Phys. 65 (1989) 2827.
[3] C. Pickering, R.T. Carline, D.J. Robbins, W.Y. Leong, S.J. Barnett, A.D. Pitt, A.G. Cullis, J. Appl. Phys. 73 (1993) 239.
[4] R.T. Carline, C. Pickering, D.J. Robbins, W.Y. Leong, A.D. Pitt, A.G. Cullis, Appl. Phys. Lett. 64 (1994) 1114.
[5] C. Pickering, R.T. Carline, J. Appl. Phys. 75 (1994) 4642. See also C. Pickering, this volume.
[6] K. Eberl, S.S. Iyer, S. Zollner, J.C. Tsang, F.K. LeGoues, Appl. Phys. Lett. 60 (1992) 3033.
[7] S. Zollner, C.M. Herzinger, J.A. Woollam, S.S. Iyer, A.P. Powell, K. Eberl, Solid State Commun. 69 (1995) 305.
[8] S. Zollner, C.M. Herzinger, J.A. Woollam, S.S. Iyer, A.P. Powell, K. Eberl, Mater. Res. Soc. Symp. Proc. 379 (1995) 205.
[9] S. Zollner, J. Appl. Phys. 78 (1995) 5209.
[10] W. Kissinger, M. Weidner, H.J. Osten, M. Eichler, Appl. Phys. Lett. 65 (1994) 3356.
[11] W. Kissinger, H.J. Osten, M. Weidner, M. Eichler, J. Appl. Phys. 79 (1996) 3016.
[12] H.J. Osten, W. Kissinger, M. Weidner, M. Eichler, Mater. Res. Soc. Symp. Proc. 379 (1995) 199.
[13] H. Lee, S.R. Kurtz, J.A. Floro, J. Strane, C.H. Seager, S.R. Lee, E.D. Jones, J.F. Nelson, T. Mayer, S.T. Picraux, Jpn. J. Appl. Phys. 34 (1995) L1340.
[14] H. Lee, J.A. Floro, J. Strane, S.R. Lee, E.D. Jones, T. Mayer, S.T. Picraux, Mater. Res. Soc. Symp. Proc. 379 (1995) 211.
[15] R. Lange, K.E. Junge, S. Zollner, S.S. Iyer, A.P. Powell, K. Eberl, J. Appl. Phys. 80 (1996) 4578.
[16] S.P. Ashburn, M.C. Öztürk, G. Harris, D.M. Maher, J. Appl. Phys. 74 (1993) 4455.
[17] K.E. Junge, R. Lange, J.M. Dolan, S. Zollner, M. Dashiell, D. Hits, B.A. Orner, J. Kolodzey, Appl. Phys. Lett. 69 (1996) 4084.
[18] M. Krishnamurthy, B.-K. Yang, W.H. Weber, Appl. Phys. Lett. 69 (1996) 2572.
[19] M.A. Berding, A. Sher, M. van Schilfgaarde, Phys. Rev. B, in press.
[20] I. Ohlídal, F. Lukeš, Opt. Acta 19 (1972) 817.
[21] J. Kolodzey, P.A. O'Neil, S. Zhang, B.A. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite, S. Ismat Shah, Appl. Phys. Lett. 67 (1995) 1865.
[22] J. Kolodzey, P.A. O'Neil, S. Zhang, B.A. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite, S. Ismat Shah, Appl. Phys. Lett 68 E (1996) 1168.
[23] B.A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R.G. Wilson, P.R. Berger, J. Kolodzey, J. Elect. Mater. 25 (1996) 297.
[24] B.A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J.W. Mayer, J. Spear, Appl. Phys. Lett. 69 (1996) 2557.
[25] R.M.A. Azzam, N.M. Bashara, Ellipsometry and Polarized Light, North Holland, Amsterdam, 1992.
[26] D.E. Aspnes, A.A. Studna, Phys. Rev. B 27 (1983) 985.
[27] L. Viña, S. Logothetidis, M. Cardona, Phys. Rev. B 30 (1984) 1979.
[28] M. Cardona, in: E. Burstein (Ed.), Atomic Structure and Properties of Solids, Academic Press, New York, 1972.
[29] S. Logothetidis, J. Appl. Phys. 65 (1989) 2416.

Size distribution of SiGeC quantum dots grown on Si(311) and Si(001) surfaces

R. Jonczyk,[a] D. A. Hits, L. V. Kulik, and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

M. Kaba and M. A. Barteau
Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716

(Received 26 November 1997; accepted 6 February 1998)

Quantum dots of $Si_{1-x-y}Ge_xC_y$ alloys with high Ge contents were grown on Si(311) and Si(001) substrates by solid source molecular beam epitaxy and were measured by atomic force microscopy. The quantum dot layers had a nominal thickness (equivalent two-dimensional) of 4 nm. The smallest quantum dots occurred for the composition $Si_{0.09}Ge_{0.9}C_{0.01}$ on Si (311), and had a 40 nm mean diameter, an 8 nm mean height, and a density of 3.3×10^{10} cm^{-2}. Quantum dots on Si(001) were larger and had less regular spacing than quantum dots on Si(311) with the same composition. Carbon decreased both the mean size and spacing of SiGe quantum dots and the ratio of size deviation to mean diameter. The presence of small uniform quantum dots for particular compositions is attributed to a reduction in the surface migration of adatoms due to decreased atomic surface diffusivity. These results suggest that quantum dot organization is controlled by composition, substrate orientation, strain, and surface diffusion. © *1998 American Vacuum Society.* [S0734-211X(98)03803-7]

I. INTRODUCTION

Recent interest in low-dimensional nanostructures is motivated by the possibility of new electrical and optical properties compared to bulk and two-dimensional (2D) layers. Sufficiently small structures are expected to exhibit quantum confinement effects at room temperature, short propagation delays, and Terabit cm^{-2} densities.[1] Ge grows on Si by the Stranski-Krastanov mode of 2D wetting for the first 3 ML, followed by three-dimensional (3D) islanding.[2] This islanding process can be utilized to achieve self-assembled quantum dot arrays. Such self-assembled or self-organized quantum dots on Si substrates are being investigated for producing nanostructures compatible with Si circuit processing.[2] The fabrication of nanostructures by self-organization during epitaxial growth avoids the limitations of conventional lithographic processing. The question is whether the resulting quantum dot arrays will have the uniformity of sizes and positions suitable for device applications.

Quantum dot formation may be affected by substrate orientation through differences in surface energy, bond density, and diffusivity. The growth of Ge quantum dots on Si has been investigated for the Si(001),[3,4] Si(110),[5] Si(111),[6] and Si(311)[2] orientations. It has been shown that Ge layers relax more readily on Si(001) than on Si(111)[4] and maintain purely 2D growth for slightly thicker layers on Si(311) than on Si(001).[2] Additionally, the growth of SiGe quantum dots on Si(001) was found to depend on surface diffusion.[7,8] Quantum dot formation has been observed in the GeC system with optical properties depending on the C fractions.[9] The formation of Ge islands on Si can be suppressed by using surfactants such as Ga and Sb (Ref. 10) and by reducing the substrate temperature to below 300 °C.[3] Adding C to SiGe alloys adds flexibility and has been studied in thin films[11] but not previously in quantum dot layers. In this article, we report on the size distribution of SiGeC quantum dots grown on Si(311) and Si(001) surfaces, and compare the effects of composition, strain, and substrate orientation.

II. EXPERIMENT

The samples were grown by solid source molecular beam epitaxy (MBE) as described elsewhere.[11] Solid thermal sources were used for Si and Ge, and a heated graphite filament was used for C. The substrate temperature was 600 °C for all layers. X-ray diffraction indicated that these conditions typically produce single crystal epitaxial layers oriented with the substrate, for thick layers. The Si atomic fraction was varied from 0 to 0.1, and the C fraction was varied from 0 to 0.01. No surfactants, such as atomic H, were used during growth. The nominal thickness was 4 nm, which is the thickness of an equivalent flat layer having the same volume as the quantum dot layer. The compositions were inferred from growth conditions calibrated by Rutherford backscattering spectrometry and by electron microprobe measurements of thicker samples grown under identical conditions.

The layers were examined by atomic force microscopy (AFM), performed with a Digital Instruments Nanoscope III using the tapping mode technique with single crystal Si cantilevers having a nominal tip radius of 10 nm.[12] All samples were taken from the center of the wafer. To further ensure that the results were not due to local variations, several AFM scans were taken across the diameter of each wafer. The size and spacing data in Table I were obtained by analyzing the features from AFM scans. The quantum dot area was taken

---

[a] Electronic mail: jonczyk@ee.udel.edu

TABLE I. Nanostructure properties including composition, orientation of Si substrates, and statistical properties of quantum dots for 2 μm×2 μm AFM scans of samples. Dev/dia is the standard deviation of the diameter distribution divided by the diameter.

Sample No.	Composition	Substr. orient.	Dot diameter (nm)	Peak wavl. (nm)	Dot density (cm^{-2})	Avg. dot height (nm)	Dev /dia
SGC191	$Si_{0.1}Ge_{0.9}$	(311)	70	141	7×10^{9}	9.1	0.86
SGC180	$Si_{0.09}Ge_{0.9}C_{0.01}$	(311)	40	64	3.4×10^{10}	7.6	0.61
SGC183	$Si_{0.1}Ge_{0.9}$	(001)	71	207	6.7×10^{9}	12.8	1.23
SGC181	$Si_{0.09}Ge_{0.9}C_{0.01}$	(001)	58	125	1.2×10^{10}	11.5	0.98
SGC204	Ge	(311)	105	231	2.5×10^{9}	16.5	0.79
SGC205	$Ge_{0.99}C_{0.01}$	(311)	95	186	3.0×10^{9}	15.6	0.74

at a vertical position located two times the root mean square (rms) surface roughness away from the mean base line 2D surface. The quantum dot diameter is calculated from the quantum dot area assuming a circle of the same area as the quantum dot.

III. RESULT AND DISCUSSION

Figure 1(a) shows a top view of SiGe quantum dots on Si (311) together with the quantum dot-area distribution. Both large coalesced quantum dots and small coherent quantum dots are present in this sample. The size distribution shows a peak for large quantum dots at about 90 nm diameter and a somewhat broader peak for small quantum dots at 35 nm. Between these two maxima quantum dots of all sizes exist at a fairly constant rate. Quantum dots grown under similar conditions but containing C are shown in Fig. 1(b). The presence of C decreased the mean size of the quantum dots, reduced the range of feature sizes, as will be shown later by Fourier analysis, and narrowed the quantum dot size distribution to only one strong peak at 40 nm quantum dot diameter. SiGeC samples grown under identical conditions on Si(001) substrates showed that C had similar effects on the quantum dot size distribution and quantum dot feature sizes, however the effect was less pronounced than for samples grown on (311). This indicates that the substrate orientation affects the quantum dot size. The addition of C to Ge quantum dots grown on (311) substrates yielded only a slight reduction in size and quantum dot spacing. We attribute the reduced effect of C on pure Ge to low substitutional incorporation of C in Ge. Statistics for all layers discussed above are given in Table I. As indicated in Table I, over the range of compositions and substrate orientations investigated, the smallest quantum dots were obtained for SiGeC compositions on Si(311). It is also worth noting that, although smaller quantum dots are expected to have a smaller deviation of sizes, C reduced the ratio of quantum dot size deviation to quantum dot diameter, indicating higher regularity in SiGe layers containing C.

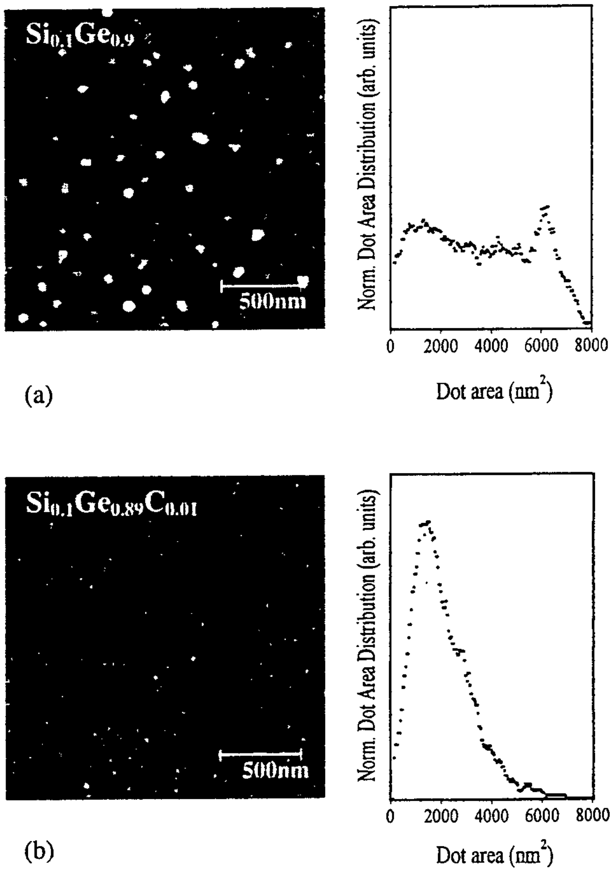


FIG. 1. (a) Atomic force microscopy scan of 2 μm×2 μm area of $Si_{0.1}Ge_{0.89}$ grown on (311) Si and quantum dot area distribution, (b) atomic force microscopy scan of 2 μm×2 μm area of $Si_{0.1}Ge_{0.89}$ $C_{0.01}$ grown on (311) Si and quantum dot area distribution. The curves were normalized by multiplying the number of equally sized quantum dots with their area and dividing the result by the scan area.

To determine the distribution of sizes, Fourier analysis was performed. The Fourier transform is proportional to the number of occurrences of the spatial wave vector, k. The spatial wavelength, $2\pi/k$, is the characteristic length associated with periodic spatial variations, including diameters, spacings, and vertical slopes of quantum dot edges. Two types of Fourier transforms were performed on the surfaces: 2D based on the two-axis (x-y) raster scans, and one dimensional (1D) based on only the x coordinate of sequential line scans. The 2D spectrum showed circular symmetry, indicating no preferred orientation on any of the substrates. The 1D Fourier intensity power spectrum, the square of the Fourier transform, is plotted versus spatial wavelength in Fig. 2, for the quantum dots of Fig. 1. The addition of C to SiGe reduced the mean value and the relative-range deviation of the

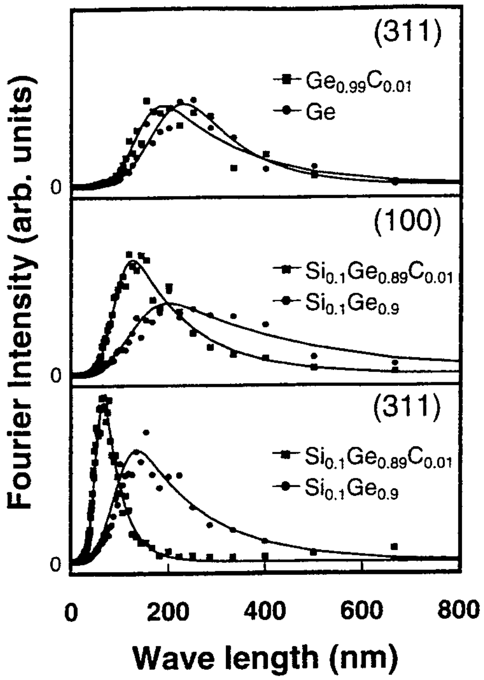


FIG. 2. Fourier intensity distribution spectrum vs quantum dot spatial wavelength for Ge, GeC, SiGe, and SiGeC quantum dots on Si(311) and Si(001), obtained from AFM scans of Fig. 1, over the 2 μm×2 μm area. Markers represent measured data, solid lines were calculated by fitting the measured data to an exponentially altered Gaussian.

quantum dot wavelengths. For the same composition the Si(311) quantum dots had a smaller size and range than the Si(001) quantum dots; possible reasons for this include differences in the energies, bond densities, and atomic diffusion coefficients of the two surfaces. It is worth noting in Fig. 2 that the SiGeC (311) sample had low intensities at wavelengths from 200 to 800 nm indicating significantly more uniformity than the other samples.

It is known that the addition of Si and C to Ge reduces the lattice mismatch to the Si substrate, thus increasing the critical thickness at which Stranski-Krastanov islanding occurs. The critical thickness t_c is proportional to the misfit[13] ϵ with $t_c \propto \epsilon^{-4}$. Adding 10% Si to Ge increases the critical thickness by about 2 ML. Substitutional C will also increase the critical thickness. We believe that a higher critical thickness results in an array of quantum dots with higher quantum dot density at the thickness immediately after the transition from 2D growth to 3D growth. This and reduced surface diffusion due to the incorporation of Si and C, which reduces the average bond length of the alloy and hence increases the diffusion barrier for surface adatoms,[14] could be responsible for the more evenly distributed smaller quantum dots observed in samples containing Si and C. The increased barrier against surface diffusion will decrease the coalescence of quantum dots into larger sizes, and thus a more closely spaced distribution of smaller quantum dots will be expected. The smaller and more regular quantum dot layers on Si(311) compared with Si(001) substrates may be due to surface-related differences including diffusivities, local strain, and critical thickness. For $Si_{0.1}Ge_{0.9}$ and $Ge_{0.99}C_{0.01}$ quantum dots under equivalent strain conditions, we found SiGe quantum dots to be smaller than GeC quantum dots, indicating that the chemical nature of the alloy affects quantum dot size in addition to strain.

IV. CONCLUSIONS

In conclusion, we have grown Ge, SiGe, GeC, and SiGeC quantum dots on (001) and (311) Si substrates. Quantum dots containing Si and C have smaller size and spacing for both (001) and (311) surfaces, and reduced deviations in size and spacing, which is crucial for applications requiring uniformity. The effect of C is stronger for alloys containing small amounts of Si, and is more prominent for alloys grown on (311) substrates. Differences between strain equivalent SiGe and GeC layers grown on (311) surfaces indicated that strain alone does not determine the geometry of quantum dot layers. These results indicate that strain, composition, and substrate orientation all play roles in quantum dot formation and provide different vehicles to control the self-assembly and self-organization of quantum dots for device applications.

ACKNOWLEDGMENTS

The authors gratefully acknowledge P. Berger, S. F. Julien, and J.-M. Lourtioz for helpful suggestions. This work was supported by the AFOSR under Contract No. F49620-95-0135, ONR under Grant No. N00014-93-10393, and DARPA under Contract No. F49620-96-C-0006.

[1] D. Goldhaber-Gordon, M. S. Montemerlo, J. C. Love, G. J. Opiteck, and J. C. Ellenbogen, Proc. IEEE **85**, 521 (1997).
[2] J. Knall and J. B. Pethica, Surf. Sci. **265**, 156 (1992).
[3] M. Hammar, F. K. LeGoues, J. Tersoff, M. C. Reuter, and R. M. Tromp, Surf. Sci. **349**, 129 (1996).
[4] P. W. Deelman, T. Thundat, and L. J. Schowalter, Appl. Surf. Sci. **104/105**, 510 (1996).
[5] M. Krishnamurthy, B.-K. Yang, J. D. Weil, and C. G. Slough, Appl. Phys. Lett. **70**, 49 (1997).
[6] P. D. Persans, P. W. Deelman, K. L. Stokes, L. J. Schowalter, A. Byrne, and T. Thundat, Appl. Phys. Lett. **70**, 472 (1997).
[7] P. O. Hansson, M. Albrecht, H. P. Strunk, E. Bauser, and J. H. Werner, Thin Solid Films **216**, 199 (1992).
[8] P. Schittenhelm, M. Gail, and G. Abstreiter, J. Cryst. Growth **157**, 260 (1995).
[9] M. Krishnamurthy, B.-K. Yang, and W. H. Weber, Appl. Phys. Lett. **69**, 2572 (1996).
[10] W. Kissinger, H. J. Osten, G. Lippert, B. Dietrich, and E. Bugiel, J. Appl. Phys. **76**, 8042 (1996).
[11] B. A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J. W. Mayer, and J. Spear, Appl. Phys. Lett. **69**, 2557 (1996).
[12] B. A. Watson and M. A. Barteau, Chem. Mater. **6**, 771 (1994).
[13] C. W. Snyder, J. F. Mansfield, and B. G. Orr, Phys. Rev. B **46**, 9551 (1992).
[14] G. Brocks, P. J. Kelly, and R. Car, Surf. Sci. **269/270**, 860 (1991).

JOURNAL OF APPLIED PHYSICS VOLUME 84, NUMBER 8 15 OCTOBER 1998

COMMUNICATIONS

Precipitation of β-SiC in $Si_{1-y}C_y$ alloys

C. Guedj,[a] M. W. Dashiell, L. Kulik,[b] and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

A. Hairie
Laboratoire d'Etudes et de Recherches sur les Matériaux (CNRS UA 6004), Institut des Sciences de La Matière et du Rayonnement, 6, Boulevard du Maréchal Juin, 14050 Caen Cedex, France

(Received 26 March 1998; accepted for publication 7 July 1998)

The infrared modes of annealed $Si_{1-y}C_y$ alloys were studied experimentally and theoretically. The alloys were grown on Si(100) substrates by solid-source molecular beam epitaxy and were characterized by Fourier transform infrared spectroscopy. At annealing temperatures above 850 °C, the localized vibrational mode of substitutional C around 605 cm^{-1} diminished in intensity while another mode due to incoherent silicon carbide precipitates appeared at 810 cm^{-1}. For lower processing temperatures, a peak around 725 cm^{-1} has been tentatively attributed to a C-rich phase, which is a precursor to SiC precipitation. Theoretical calculations based on the anharmonic Keating model predict that small (1 nm) 3C–SiC *coherent* precipitates may actually produce a mode at 725 cm^{-1}. This mode occurs if the bonds gradually vary in length between the C-rich region and the host lattice. On the other hand, if the bonds are abruptly distorted at the edges of the precipitate, it becomes elastically isolated from the host lattice, and the 810 cm^{-1} mode appears. This study yields a picture of the thermal stability of dilute SiC alloys, which is important for the high-temperature processing steps necessary for device applications. Moreover, the coherent precipitation may provide a controllable way to form self-assembled 3C–SiC quantum dots into silicon germanium carbon alloys. © *1998 American Institute of Physics.*[S0021-8979(98)08219-X]

Carbon incorporation into silicon-based microelectronics and optoelectronics provides device opportunities through band-gap and strain engineering,[1] and diffusion control. Substitutional carbon modifies the alloy band structure and the band offsets for Si-based heterostructures.[2,3] In the dilute limit, the lowest-energy configuration for C atoms is the occupation of substitutional sites,[4] where the carbon atoms are tetrahedrally bound to Si atoms. The difference in lattice parameters and covalent radii between silicon and diamond induce high bond distortions, which results in a modification of the phonon spectrum.[5] In the limit of the highly diluted $Si_{1-y}C_y$ alloy, substitutional carbon gives a signature situated around 605 cm^{-1}. This localized vibrational mode can be observed either by absorption spectroscopy[6] or by Raman spectroscopy.[5,7,8] After thermal annealings at high temperatures (>900 °C), a broad peak around 810 cm^{-1} has been observed by several groups.[6,9,10,11] This phenomenon is usually attributed to silicon carbide precipitation.[11] This result is naturally expected since the solubility of C in Si is 3×10^{17} cm^{-3} at the melting point[11] and stoichiometric SiC is the only thermodynamically stable structure in the Si–C system.[12] An additional vibrational mode situated around 725 cm^{-1} has been recently observed[13] by infrared absorption spectroscopy for intermediate growth temperatures (~650 °C). This peak is consistent with a Si–C transitional phase, intermediate between the loss of substitutional C sites on the silicon lattice and the formation of SiC-relaxed precipitates. Here, we use a theoretical approach to demonstrate that the *coherent* precipitation of 3C–SiC in Si produces a vibrational mode situated between 700 and 740 cm^{-1} because of the strain-affected chemical bonds around C. The phonon mode around 810 cm^{-1} is obtained only in the case of *incoherent* precipitates. The latter case is characterized by the absence of crystalline continuity between the precipitates and the surrounding matrix, and the interface strain between 3C–SiC and Si may be accommodated by defects.

Our theoretical approach is based upon a valence force field model derived from Keating[14] and taking into account the effects of carbon anharmonicity. The interactions between atoms have been modeled with an interatomic potential similar to the one of Rücker *et al.*[5] with the exception of adjusting our force coefficients to yield the correct lattice parameters and phonon modes of Si, diamond (in the Fd3m structure) and 3C–SiC at 300 K. This choice is justified by the very high precision and reliability of these experimental data, which can be obtained by x-ray diffraction, Raman spectroscopy, and absorption spectroscopy at room temperature. In addition, our set of parameters enables the computa-

---

[a] Corresponding author: electronic mail: guedj@ee.udel.edu

[b] On leave from the Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow region, Russia.

© 1998 American Institute of Physics

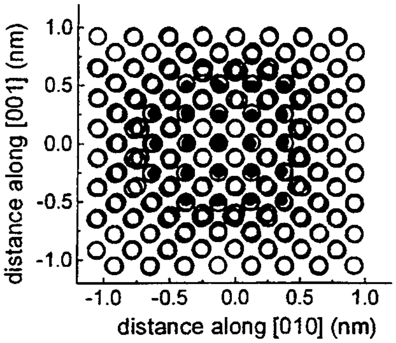


FIG. 1. View along [100] of a 3C–SiC coherent precipitate coherently embedded into silicon and containing 50 carbon atoms. Molecular dynamics computation is used to relax the 512 atom supercell isotropically. The solid and open spots stand for C and Si, respectively. A modified Keating model, which takes into account the anharmonicity of C, is used to model the atomic interactions.

tion of the LO(Γ) mode from 3C–SiC, which cannot be obtained from the parameters given by Rücker *et al.*[5] to the best of our knowledge. Our model is, therefore, ideally suited to a precise computation of lattice parameters and phonon spectra of tetrahedrally coordinated $Si_{1-y}C_y$ alloys. The molecular dynamics relaxation is calculated from a 512 atom supercell by time increments of 3 fs, until the total potential energy reaches a stable minimum. We have computed both isotropic relaxations, where no external pressure is applied to the computational box, and pseudomorphic relaxations, to simulate a fully strained layer on silicon. In a further stage, the local phonon density around carbon is computed, using the recursion method detailed in Ref. 15. This algorithm enables the computation of the local phonon density around chosen atoms. The calculated spectra can be compared with absorption spectroscopy and Raman spectroscopy results after application of proper selection rules. Numerous random and ordered atomic configurations have been tested, in order to simulate the precipitation of 3C–SiC into silicon.

The application of these theoretical tools gives the following results. In the case of a 3C–SiC precipitate containing 50 carbon atoms and coherently embedded into a silicon matrix, the atomic positions after isotropic relaxation are depicted in Fig. 1. The high bond distortions of this quantum dot are clearly evident, especially around the precipitate. From that picture, it is clear that the interface energy between the matrix and the precipitate is important, in agreement with Taylor *et al.*[16] Silicon carbide is more rigid than silicon, therefore, the interface between the silicon and 3C–SiC is very abrupt, and the distortion occurs across few atomic distances. The statistical distribution of Si–C bonds is represented in Fig. 2. The distribution is broadened and asymmetrically distorted by the silicon-related strain. The Gaussian fit is centered around 0.208 nm, which is between 3C–SiC (0.188 nm) and Si (0.235 nm). The more distorted bonds can reach up to 0.235 nm at the interface between the matrix and the precipitate. This local departure from perfect crystalline periodicity gives an intense phonon mode, illustrated in Fig. 3(b). This spectrum is roughly composed of the classical localized vibrational mode around 605 cm^{-1}, a small peak around 660 cm^{-1}, and a well-defined mode at about 725 cm^{-1}. Silicon–carbon bonds situated about one spherical monolayer inside the outermost frontier of the precipitate give the most intense phonon modes around 725 cm^{-1}. Curve (c) in Fig. 3 represents the local phonon density spectra averaged to all carbon atoms in the supercell. The 605 cm^{-1} localized vibrational mode is spread toward higher wavenumbers because of the high tensile strain, and the 725 cm^{-1} peak is present. These simulations are in reasonable agreement with the experimental absorption spectroscopy spectrum depicted in Fig. 3(a). The position of the mode depends upon the size of the 3C–SiC quantum dot and the strain around the precipitate. If the supercell is pseudomorphic to silicon, then the peak is situated around 560 cm^{-1}. The potential energy of the system is well fitted by a sigmoidal (Boltzman) increasing function of the precipitate size, and the phonon peak shifts toward higher energies

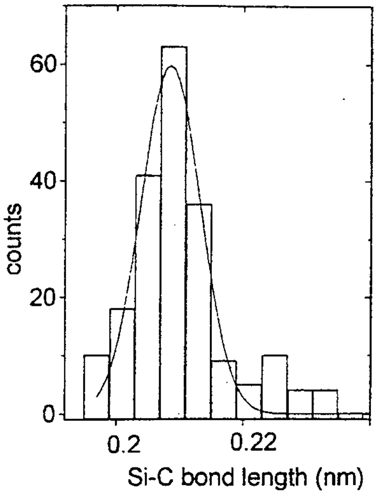


FIG. 2. Statistical distribution of Si–C bonds in the 3C–SiC precipitate of Fig. 1. A Gaussian fit of the distribution is centered around 0.208 nm. The calculated bond length can reach up to 0.235 nm. For comparison, the Si–C bond length in pure 3C–SiC is 0.188 nm.

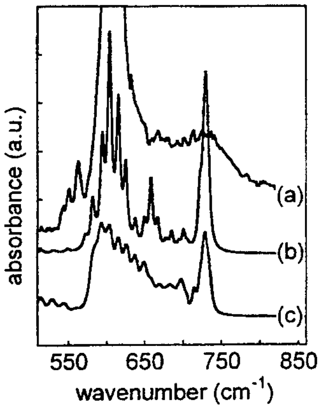


FIG. 3. Comparison between the experimental infrared absorption spectrum and the local phonon density calculated by the recursion method, for the 3C–SiC precipitate illustrated in Fig. 1. The experimental curve (a) is obtained with a $Si_{1-y}C_y$ sample grown by molecular beam epitaxy at 650 °C. The simulated curve (b) is calculated around a single carbon atom situated at the interface between the precipitate and its surrounding matrix, while (c) is the phonon spectrum averaged over all carbon atoms in the precipitate.

when the size increases or when the precipitate becomes incoherent to its matrix. The mode observed at 725 cm^{-1} is best matched in energy by the phonon spectrum of a spherical quantum dot containing about 50 carbon atoms and coherently embedded into an isotropically relaxed matrix.

Now, we propose the following interpretation for the physical behavior of substitutional carbon into silicon (or silicon germanium) alloys during thermal annealing. For rapid thermal annealings above 1000 °C during a few seconds or at temperatures higher than 850 °C for several hours, the formation of relaxed silicon carbide precipitates seems to be unanimously accepted.[6,9–11] This process is attributed to thermally activated carbon diffusion, followed by precipitation. The interface between 3C–SiC and the matrix is defective, and the nanoprecipitates are relaxed. Our calculations show that if the precipitate were coherent, then the Si–C bond length would increase up to 25% at the interface. Temperature is linked to atomic agitation, therefore, structural defects are very likely to occur at the interface for high thermal atomic vibrations. Various types of defects may be involved, either with carbon, oxygen, or other complexes. For intermediate processing temperatures (between 600 and 850 °C approximately), we suggest that substitutional carbon should have an increasing probability to jump into interstitial sites, for example, by forming a Si–C interstitial complex. It may, thereafter, produce a vacancy. Then, it could rapidly diffuse to form small-sized coherent precipitates, surrounded by a vacancy-rich matrix. The formation of such quantum dots may also depend on surface effects during the metastable growth. The additional presence of various defects might not be completely excluded, and thorough investigations have to be performed to further evaluate their nature and their density. The strain may be roughly maintained if the precipitates are coherent, because the crystalline periodicity is not completely disrupted. For obvious reasons, local elastic energy and Si–C bonds maximal distortions cannot increase indefinitely, therefore, such coherent precipitates should be highly metastable. A simulated size of ~1 nm best matches our infrared absorption experimental results, therefore, the critical radius for loss of coherency of β-SiC in Si should be superior to 1 nm for our growth conditions. Similar results have been observed with silicon germanium carbon alloys. This mechanism of coherent precipitation may provide a controllable way to form self-assembled quantum dots of ~1 nm size, which can be observed by absorption spectroscopy.

In summary, we have performed molecular dynamics simulations using a valence force field model to simulate the silicon carbide precipitation into silicon. The recently measured[13] phonon mode situated around 725 cm^{-1} is computed in the case of a *coherent* 3C–SiC precipitate isotropically relaxed. This picture is consistent with ~1 nm β-SiC quantum dots coherently embedded into a vacancy-rich matrix. The phonon mode around 810 cm^{-1} is obtained only if the silicon carbide nanoprecipitates are *incoherent*. This mechanism of coherent precipitation may provide a controllable way to form self-assembled β-SiC quantum dots in silicon germanium carbon alloys.

The authors gratefully acknowledge support from ARO under Grant No. DAAH04-95-1-0625, and DARPA Contract No. F49620-96-C-0006.

[1] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, Appl. Phys. Lett. **60**, 3033 (1992).
[2] K. Brunner, K. Eberl, and W. Winter, Phys. Rev. Lett. **76**, 303 (1996).
[3] M. Kim and H. J. Osten, Appl. Phys. Lett. **70**, 2702 (1997).
[4] J. Tersoff, Phys. Rev. Lett. **64**, 1757 (1990).
[5] H. Rücker, M. Methfessel, B. Dietrich, K. Pressel, and H. J. Osten, Phys. Rev. B **53**, 1302 (1996).
[6] J. W. Strane, H. J. Stein, S. R. Lee, S. T. Picraux, J. K. Watanabe, and J. W. Mayer, J. Appl. Phys. **76**, 3656 (1994).
[7] C. Guedj, X. Portier, A. Hairie, D. Bouchier, G. Calvarin, and B. Piriou, Thin Solid Films **294**, 129 (1997).
[8] M. Meléndez-Lira, J. Meléndez, K. M. Kramer, M. O. Thompson, N. Cave, R. Liu, J. W. Christiansen, N. D. Theodore, and J. J. Candelaria, J. Appl. Phys. **82**, 4246 (1997).
[9] H. J. Osten, M. Kim, K. Pressel, and P. Zaumseil, J. Appl. Phys. **80**, 6711 (1997).
[10] P. Warren, J. Mi, F. Overney, and M. Dutoit, J. Cryst. Growth **157**, 414 (1995).
[11] A. R. Bean and R. C. Newman, J. Phys. Chem. Solids **33**, 255 (1971).
[12] R. W. Olesinski and G. J. Abbaschian, Bull. Alloy Phase Diagrams **5**, 5 (1984).
[13] M. W. Dashiell, L. V. Kulik, D. Hits, J. Kolodzey, and G. Watson, Appl. Phys. Lett. **72**, 833 (1998).
[14] P. N. Keating, Phys. Rev. **145**, 637 (1966).
[15] A. Hairie, F. Hairie, G. Nouet, E. Paumier, and A. P. Sutton, in *Polycristalline Semiconductors III—Physics and Technology*, edited by H. P. Strunk, J. H. Werner, B. Fortin, and O. Bonnaud (Trans Tech Publications Ltd., Aedermannsdorf, Switzerland, 1994), Vol. 37-38, p. 91.
[16] W. J. Taylor, T. Y. Tan, and U. Gösele, Appl. Phys. Lett. **62**, 3336 (1993).

The Effects of Composition and Doping on the Response of GeC–Si Photodiodes

James Kolodzey, *Senior Member, IEEE,* Olivier Gauthier-Lafaye, Sébastien Sauvage, Jean-Luc Perrossier,
Philippe Boucaud, Francois H. Julien, Jean-Michel Lourtioz, *Senior Member, IEEE,* Fen Chen,
Bradley A. Orner, Kristofer Roe, Cyril Guedj, R. G. Wilson, and Jennifer Spear

Abstract—**The spectral responses of a series of heterojunction diodes of p-type $Ge_{1-y}C_y$ on n-type Si (100) substrates were measured by Fourier transform infrared (IR) spectroscopy. Alloy layers 0.5 μm thick were grown by molecular beam epitaxy at a substrate temperature of 400 °C and were doped p-type with different B concentrations. With increasing C content, the diode dark current decreased, and the optical absorption band edge shifted toward higher energy by 70 meV for 0.12 atomic persent of C. The increase in energy was attributed to the composition dependence of the bandgap rather than to strain relaxation, because the GeC layers were nearly relaxed with the same strain. The photoresponsivity was 0.07 A/W at a wavelength of 1.55 μm, and 0.2 A/W at a wavelength of 1.3 μm. These measurements show that GeC photodetectors have good properties and reasonable response at technologically important near-IR wavelengths and can be fabricated by heteroepitaxy for compatibility with Si integrated circuits.**

Index Terms— **Epitaxial growth, germanium alloys, infrared detectors, infrared spectroscopy, optical measurements, photodiodes, semiconductor heterojunctions.**

I. INTRODUCTION

INFRARED (IR) photodetectors (PD's) for the optical fiber communication wavelengths of 1.3 and 1.55 μm are typically made of Ge or InGaAs [1]. Epitaxial IR detectors on Si substrates are attractive for circuit integration, provided that the detector properties and the reliability are adequate [2]. The 4% lattice mismatch of epitaxial Ge on Si produces strain that relaxes by dislocation and defect formation in thick layers [1]. The addition of C to Ge reduces the strain thus limiting the formation of dislocations, and also reduces the diffusion of dopants [3]. The effects of C on device properties are not yet well understood. The optical and electrical characteristics of GeC photodiodes are therefore interesting both fundamentally and technologically. We report here on measurements of the photoresponse of GeC–Si heterojunction photodiodes on Si substrates, to determine if the layers are suitable for device applications.

II. EXPERIMENTAL APPROACH

A. Device Fabrication

The p-doped GeC layers were grown by solid source molecular beam epitaxy (MBE) at 400 °C using methods described elsewhere [4], [5]. The substrates were n-type Si (100) with a doping concentration of 3×10^{15} cm^{-3}. The Ge source was a thermal effusion cell operating from 1310 °C to 1325 °C with a pyrolytic boron nitride (pBN) crucible. The C source was a pyrolytic graphite filament heated by currents up to 48 A. The p-type dopant source was a high temperature effusion cell containing elemental B in a crucible of pyrolytic graphite in a W metal jacket, operating from 1450 °C to 1650 °C. The alloy growth rate was 1.5 nm/min. No surfactants such as atomic H were used during growth, and no changes in growth rate with doping or composition were observed. The layers were 560 nm thick, except for one sample that was 520 nm thick (SGC-84 in Table I).

X-ray diffraction indicated that the epitaxial layers were single-crystals oriented to the substrate. Measurements of the symmetric and asymmetric X-ray reflections indicated that the strain was near 10^{-3} so that the alloy layers were nearly relaxed. In Table I, the strain is given by the difference between the perpendicular and parallel lattice constants divided by the relaxed lattice constant. The substitutional C fractions given in Table I were obtained from the effective lattice constants of fully relaxed alloys deduced by analyzing the X-ray diffraction measurements [6]. A linear relation was assumed for the dependence of the alloy lattice constant on composition for the $Ge_{1-y}C_y$. It is possible that the total C fraction may be higher than the substitutional values quoted here due to bowing in the relation between lattice constant and composition. RBS ion channeling performed on similar samples with higher C contents indicated that at least 80% of the C occupied substitutional lattice sites [6]. The material quality of the layers was reported previously, and indicated that the layer/substrate interface had defects including twins [7].

The B doping concentration was measured using secondary ion mass spectrometry (SIMS). SIMS profiling showed that

---

Manuscript received March 2, 1998; revised September 18, 1998.

J. Kolodzey, F. Chen, B. A. Orner, K. Roe, and C. Guedj are with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA.

O. Gauthier-Lafaye, S. Sauvage, J.-L. Perrossier, P. Boucaud, F. H. Julien, and J.-M. Lourtioz are with the Institut d'Electronique Fondamentale, Université Paris-Sud—Bât. 220, URA CNRS 22, 91405 Orsay, France.

R. G. Wilson is at 25234 B Steinbeck Avenue, Stevenson Ranch, CA 91381 USA.

J. Spear is with the Philips Electronic Instruments Company, Tempe, AZ 85283 USA.

This work was supported by the French CNRS, by the Army Research Office under Grant DAAH04-95-1-0625 and Grant AASERT DAAG55-97-1-0249, by the Defence Advanced Research Projects Agency under Contract F49620-96-C-0006, and by the Office of Naval Research under Grant N00014-93-1-0393.

Publisher Item Identifier S 1077-260X(98)09127-8.

1077–260X/98$10.00 © 1998 IEEE

TABLE I

PROPERTIES OF THE $Ge_{1-y}C_y$ LAYERS USED IN THE p-N HETEROJUNCTION PD'S INCLUDING COMPOSITION, B DOPING LEVEL, FORWARD CURRENT IDEALITY FACTOR, BANDGAP ENERGY, PHONON ENERGY AND STRAIN. THE n-TYPE Si (001) SUBSTRATES HAD A DOPING LEVEL OF 3×10^{15} CM^{-3}. THE HOLE CARRIER CONCENTRATIONS ARE LISTED BENEATH THE B IMPURITY CONCENTRATIONS. THE LAYER THICKNESS WAS 560 NM FOR ALL SAMPLES EXCEPT SGC-84, WHICH WAS 520 NM THICK. THE GE DIODE IS A COMMERCIAL BULK GE HOMOJUNCTION AS DESCRIBED IN THE TEXT.

Sample	C content (atomic %)	B conc. hole conc. (cm^{-3})	Forward ideality factor	E_g (eV)	E_{ph} (meV)	Strain
Ge diode	0		-	0.639	10	-
SGC102	0	B: $1.2x\ 10^{17}$ p: $7.5x10^{17}$	1.04	0.657	14	2.8×10^{-3}
SGC103	0	B: $1.2x\ 10^{19}$ p: $1.8x10^{19}$	1.84	0.634	17	2.4×10^{-3}
SGC80	0.06%	B: $3x\ 10^{18}$ p: $2.5x10^{18}$	1.22	0.731	11	2.7×10^{-3}
SGC81	0.08%	B: $4x\ 10^{17}$ p: $6.8x10^{17}$	1.09	0.741	11	3.1×10^{-3}
SGC82	0.11%	B: $3x\ 10^{19}$ p: $1.1x10^{19}$	1.08	0.715	12	3.2×10^{-3}
SGC84	0.12%	B: $5x\ 10^{18}$ p: $1.9x10^{18}$	1.01	0.728	11	2.6×10^{-3}

the doping was constant versus depth. Measurements across the wafer surface indicated that the doping level was uniform within 10% over a lateral distance of 1 cm. Table I gives the physical properties of the samples. For the same growth conditions, there is a slight trend of higher B concentration with C fraction, possibly due to an increase in the sticking coefficient of B with C, or perhaps to an artifact due to composition-dependent yields with SIMS. Hall effect measurements yielded the hole carrier concentrations (p) given in Table I that were reasonably close to the B doping concentration except for sample SGC-102, perhaps due to experimental error.

Mesa diodes with a junction area of 0.2 mm^2 and a light sensitive area (not shaded by the contacts) of 0.182 mm^2 were fabricated using photolithography and wet chemical etching [8]. Electrical contacts of Ti–Au metal were thermally evaporated to a thickness of 60 nm/300 nm, respectively, as shown in the inset to Fig. 1.

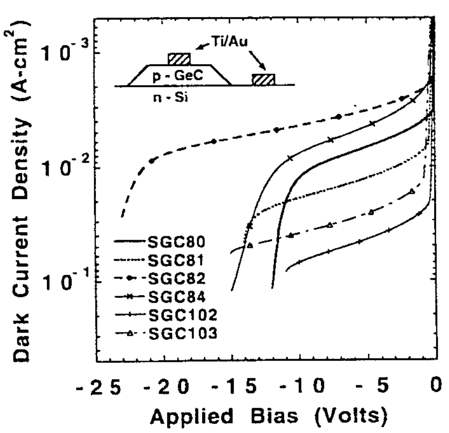


Fig. 1. Dark current density versus reverse bias voltage of p-N type $Ge_{1-y}C_y$–Si heterojunction diodes on Si (001) substrates with different composition and doping, given in Table I. The reverse leakage currents decreased significantly with C fraction and with doping concentration. The inset shows the diode structure.

B. Electrical Measurements

Current versus voltage (I–V) measurements at room temperature showed rectifying characteristics, with reverse breakdown voltages ranging from 11 to 24 V, as in Fig. 1. The reverse leakage dark current decreased significantly with increasing C content and with increasing hole concentration. In photodiode applications, a low dark current produces less noise and improves the detectivity [1]. The decrease in dark current with C content was attributed to reductions in the bulk and surface recombination rates, possibly due to: (1) a decrease in the intrinsic carrier concentration n_i in the GeC layer, (2) a reduced dislocation density due to strain compensation, and (3) changes in the energy bandgap and band offsets. The diode p-n junctions were exposed on the mesa sidewalls, and an extra processing step of passivating the junctions with SiO_2, for example, may help reduce the dark current.

For completeness, the forward bias ideality factors are included in Table I, although it is not our intent to report here on the forward bias characteristics of the diodes. Ideality factors near 1 imply diffusion current, and ideality factors near 2 imply recombination current. Due to the high series resistance of the diodes, the values given here were obtained at small forward bias and may have significant error. A more highly doped substrate and different contact metals may help reduce the series resistance.

C. Optical Measurements

The spectral response of the photodiodes was measured at room temperature using a Biorad FTS-60A Fourier transform IR spectrometer (FTIR) over the spectral range from 6000 to 10 000 cm^{-1}, using a quartz beam splitter and a glowbar light source. FTIR resolutions of 16 to 32 cm^{-1} were used for these measurements. The FTIR spectrometer was equipped with a UMA 500 microscope, and the photodiodes were inserted into the microscope beam path, using needle probes for electrical contact. To reduce the contribution of the Si substrate to the diode photocurrent, a high-resistivity Si wafer, polished on both sides, was placed in the FTIR beam path to remove photon energies above 1.1 eV. A low impedance I–V amplifier (Hewlett-Packard 4140B), which was ac-coupled to the FTIR electronics, collected the diode photocurrent. The ac-coupling of the measured signal yields the photocurrent, with the dc dark current removed. The diode photocurrent was referenced to the measured reflectivity of a thick Au metal pad located on each photodiode sample. The Au reflectivity was measured using the mercury cadmium telluride (MCT) detector of the FTIR microscope, which was calibrated by a pyroelectric (DTGS) detector having flat response versus wavelength. This procedure corrected the possible variations in instrument alignment and drift. The error was below 3 meV, based on consecutive measurements of the same diode. A commercial bulk Ge photodetector with known calibration was measured as a reference for the absolute responsivity. The commercial photodiode was a homojunction of Ge with an area of 8×10^{-3} cm^2, and a dark current of 2.5×10^{-3} A·cm^{-2} at a bias of −3 V.

Fig. 2 shows the photoresponsivity versus photon energy for diodes with different alloy compositions and with hole concentrations ranging from 10^{17} to 10^{18} cm^{-3} (see Table I). The diodes were biased at −3 V. As seen, the responsivity values are respectable, considering that the GeC active layers are relatively thin. The strong responsivity above 0.8 eV was attributed to absorption by the direct energy valley of the alloy [9]. In comparison, the responsivity of a thick bulk Ge photodiode followed the same energy dependence as that of the Ge/Si diodes near the indirect bandgap of Ge ($E_g = 0.66$ eV), reaching a maximum of 0.7 A/W at 0.8 eV and then slightly decreasing due to surface absorption. At energies near the fundamental bandgap of Ge, the onset of photoresponse regularly increased in energy with C concentration. A comparison of diodes having the same doping level yielded an increase in the absorption energy of ≈70 meV due to the presence of 0.12 atomic % C. Since the GeC alloy layers were thick and nearly relaxed with similar residual strain, the shift in absorption energy was attributed to the alloying effect of C, rather than to strain compensation. The Si layer is relatively transparent at these wavelengths, and the photoresponse was attributed fully to the GeC layer. At zero bias, the diodes showed a similar increase in absorption energy with C, but the responsivity magnitude was about half that at −3-V bias at all wavelengths. The lower photoresponse was attributed to the narrower depletion width at zero bias, and to less efficient charge collection at the lower electric field. In these junction photodiodes, the photogenerated charge is collected both within the depletion region and within a diffusion length of the depletion edge.

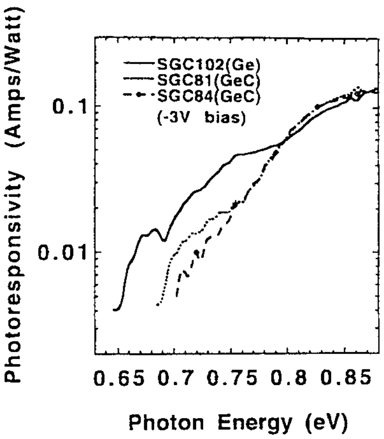


Fig. 2. Spectral responsivity of $Ge_{1-y}C_y$–Si p-N heterojunction diodes with different C fractions at −3-V reverse bias. All $Ge_{1-y}C_y$ layers had nearly the same strain and thickness (as in Table I). The samples shown here had hole concentrations from 10^{17} to 10^{18} cm^{-3}. The blue-shifted absorption was attributed to the alloying effect of C as indicated by the bandgaps in Table I. For a given C content, the bandgap energies decreased with doping concentration, as explained in the text. At zero bias, the spectral responsivity of all the diodes was reduced to about half of the magnitude at −3 V.

The photoresponsivities of two samples with hole concentrations near 10^{19} cm^{-3} (SGC-103 and SGC-82) were shifted to slightly lower energies relative to the more lightly doped samples. The explanation may be found in bandgap renormalization, which decreases the absorption edge in semiconductors, and to impurity band absorption [10]. The highly doped samples have a lower responsivity at all energies, attributed to their smaller depletion widths and diffusion lengths.

III. Analysis and Discussion

The results of Fig. 2 were exploited to estimate the indirect bandgap energy of the GeC absorbing layer of the photodiodes. For diodes with an optical absorbing width W that is less than the absorption depth $1/\alpha$ (the reciprocal of the optical absorption coefficient), the photocurrent density is approximately $J_{\mathrm{ph}} = q(1 - R)P\alpha W/h\upsilon$, where q is the magnitude of the electron charge, R the surface optical reflectivity, P the incident optical power density, and $h\upsilon$ the photon energy. For energy bands that are indirect in k space, the absorption coefficient follows the Macfarlane–Roberts expression [9], [11] for single phonon emission and absorption, with energy dependence given by

$$\alpha = A\left[\frac{(hv - E_g - E_{\mathrm{ph}})^2}{1 - e^{-E_{\mathrm{ph}}/kT}} + \frac{(hv - E_g + E_{\mathrm{ph}})^2}{e^{E_{\mathrm{ph}}/kT} - 1}\right]$$

which holds for photon energies such that

$$hv > E_g + E_{\mathrm{ph}}$$

with phonon energy E_{ph}, thermal energy kT, and where A is a constant.

In principle, a plot of the square root of photocurrent versus hv can yield two linear slopes that extrapolate to zero absorption giving $E_g \pm E_{\mathrm{ph}}$. Due to the weak photocurrent at energies near the bandgap, however, this procedure can give significant error. As a more comprehensive approach, we curve-fit the full Macfarlane-Roberts expression with the bandgap energy and the phonon energy as fitting parameters. The results of the fits for E_g and E_{ph} are reported in Table I. The values obtained for E_{ph} are consistent with the results of high-resolution optical absorption studies in Ge indicating that acoustic phonons are more strongly involved than optical phonons [11].

The increase in the absorption edge with the addition of C is consistent with previous studies of Ge-rich alloys. An increase of 63 meV per at. % C was previously obtained from optical transmission measurements of unstrained GeC alloys [4]–[6]. An increase of 43 meV per at. % C for the E_1 critical point near the L minimum was obtained by spectroscopic ellipsometry after correcting for strain [12]. Assuming a deformation potential of 10 eV [10], the variation in strain between the samples would account for at most 8 meV of the observed bandgap differences. The reduction in the diode dark current by about a factor of 10 is consistent with the observed 70-meV increase in bandgap, because the leakage caused by recombination is proportional to the square of the intrinsic carrier concentration n_i^2, which decreases exponentially with bandgap energy. It is also possible, however, that the lower leakage is due to a reduction in the dislocation density by C-induced strain compensation. Fig. 3 summarizes the dependence of the bandgap on composition and doping obtained from the curve fittings of the photoresponses. The reason for the initial strong increase in bandgap for small C fractions is not clear, but may indicate a strong bowing in the relation of bandgap to composition, or to the effects of carrier transport on the photoresponse.

Included in Fig. 3 is the result of theoretical calculations of E_g reported previously [13] based on the virtual crystal approximation (VCA) using the linear combination of atomic orbitals (LCAO) method. This theory predicts an increasing bandgap with C fraction, although the VCA approach has the limitations of not accounting for strain localized about the C atom, nor to energy band broadening from alloy disorder [14]. The theoretical curve was included for comparison; not as a claim for the validity of the VCA approximation. For Ge-rich alloys, the increase in bandgap with C is opposite to the decreases in bandgap for Si-rich alloys reported from measurements [15], and from predictions based on first principles calculations [16]. These opposing effects of C with composition may involve the differences in the behavior of the L energy band minima for Ge versus the X minima for Si.

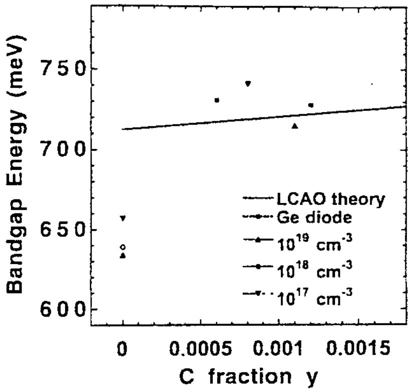


Fig. 3. The dependence of bandgap energy (E_g) on C faction and doping level. The solid symbols give the doping concentrations measured by SIMS. The data points give E_g for the samples of Table I, obtained by fitting the measured diode photocurrents to the Macfarlane–Roberts expression for phonon-assisted absorption. The variations in E_g at particular C fractions are not due to experimental uncertainty, but show the decrease of E_g with doping. For comparison purposes only, the solid line gives the theoretical E_g calculated by the LCAO method using the virtual crystal approximation [13], with a slope of 25.7 meV per at. % C. For brevity, the doping concentrations listed in the legend are approximate; the precise values are given in Table I.

The Macfarlane–Roberts expression for α accurately fit the diode photoresponses for energies above the indirect bandgap of Ge. The bandgap of diode SGC102 diode was in good agreement with the commonly accepted value for Ge. The bandgap measured for the commercial Ge diode was in-between that obtained for the Ge–Si diodes SGC102 and SGC103. Although the doping of the commercial Ge diode was not known, a possible reason for its bandgap could be that it has a doping level in-between that of the two SGC Ge diodes.

IV. CONCLUSION

In summary, the electrical and optical measurements showed that $Ge_{1-y}C_y$–Si heterojunction photodiodes have a responsivity that is attractive for practical applications. The photodiodes were fabricated by heteroepitaxy using MBE at low temperatures for compatibility with Si integrated circuit technology. Alloying with C increased the bandgap energy of $Ge_{1-y}C_y$, and was attributed to composition effects rather than to strain compensation because the layers were nearly relaxed. The variable bandgap can be useful for tuning the desired operating wavelengths of photodetectors and for reducing unwanted signals without sacrificing detector performance at important optical fiber communication wavelengths such as 1.55 and 1.3 μm. The reduction in dark current by adding C to Ge can improve the signal-to-noise ratio of optical receivers.

ACKNOWLEDGMENT

Special thanks to J. W. Mayer, T. Laursen, and D. Hits for assistance with the composition measurements, to H. Feng and P. R. Berger for sample preparation, and to A. Misera, M. Dashiell and T. Troeger for the Hall effect measurements.

Appendix 39

REFERENCES

[1] J. C. Bean, "Silicon-based semiconductor heterostructures: Column IV bandgap engineering," in *Proc. IEEE*, vol. 80, pp. 571–587, 1992.

[2] S. Murtaza, R. Mayer, M. Rashed, D. Kinosky, C. Mazier, S. Banerjee, A. Tasch, J. C. Campbell, J. C. Bean, and L. J. Peticolas, "Room temperature electroabsorption in a GeSi PIN photodiode," *IEEE Trans. Electron Devices*, vol. 41, pp. 2297–2300, 1994.

[3] H. Feng, M. Dashiell, B. A. Orner, J. Kolodzey, P. R. Berger, M. H. Ervin, and R. T. Lareau, "Boron diffusion in SiGeC alloys grown on a silicon substrate," presented at the 39th Electronic Materials Conf., Fort Collins, CO, June 25–27, 1997.

[4] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. I. Shah, "Growth of germanium-carbon alloys on silicon substrates by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 67, pp. 1865–1867, 1995.

[5] B. A. Orner, A. Khan, D. Hits, F. Chen, K. Roe, J. Pickett, X. Shao, R. G. Wilson, P. R. Berger, and J. Kolodzey, "Optical properties of GeC alloys," *J. Electron. Mater.*, vol. 25, pp. 297–300, 1996.

[6] B. A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J. W. Mayer, and J. Spear, "Bandgap of Ge-rich SiGeC alloys," *Appl. Phys. Lett.*, vol. 69, pp. 2557–2559, 1996.

[7] F. Chen, R. T. Troeger, K. Roe, M. D. Dashiell, R. Jonczyk, D. S. Holmes, R. G. Wilson, and J. Kolodzey, "Electrical properties of SiGeC and GeC alloys," *J. Electron. Mater.*, vol. 26, pp. 1371–1375, 1997.

[8] F. Chen, B. A. Orner, D. Guerin, A. Khan, P. R. Berger, S. I. Shah, and J. Kolodzey, "Current transport characteristics of SiGeC/Si heterojunction diodes," *IEEE Electron Device Lett.*, vol. 17, pp. 589–591, 1996.

[9] R. Braunstein, A. R. Moore, and F. Herman, "Intrinsic optical absorption in germanium-silicon alloys," *Phys. Rev.*, vol. 109, pp. 695–710, 1958.

[10] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors*. Berlin, Germany: Springer-Verlag, 1996.

[11] G. G. Macfarlane, T. P. McLean, J. E. Quarrington, and V. Roberts, "The fine structure in the absorption-edge spectrum of Ge," *Phys. Rev.*, vol. 108, pp. 1377–1383, 1957.

[12] S. Zollner, "Theory of optical interband transitions in strained SiC grown pseudomorphically on Si (001)," *J. Appl. Phys.*, vol. 78, pp. 5209–5212, 1995.

[13] B. A. Orner and J. Kolodzey, "SiGeC alloy band structures by linear combination of atomic orbitals," *J. Appl. Phys.*, vol. 81, pp. 6773–6780, 1997.

[14] C. G. Van de Walle and R. M. Martin, "Theoretical calculations of heterojunction discontinuities in the Si/Ge system," *Phys. Rev.*, vol. B34, pp. 5621–5627, 1986.

[15] A. S. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewalt, "Defect-free band-edge photoluminescence and band gap measurements of pseudomorphic SiGeC alloy layers on Si (100)," *Appl. Phys. Lett.*, vol. 67, pp. 3915–3917, 1995.

[16] M. A. Berding, A. Sher, and M. van Schilfgaarde, "Group IV semiconductor compounds," *Phys. Rev.*, vol. B56, pp. 3885–3891, 1997.

James Kolodzey (M'74–SM'90) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1986, for research on amorphous SiGe alloys.

He worked at IBM Corporation, and at Cray Research. From 1986 to 1989, he served as Assistant Professor of Electrical Engineering at the University of Illinois at Urbana—Champaign where he established a laboratory for high frequency device measurements at cryogenic temperatures. In 1987 he worked on molecular beam epitaxy with Dr. A. Y. Cho at AT&T Bell Laboratories. In 1990, he worked on SiC and SiGeC alloys with F. Koch and R. Schwarz at the Technical University of Munich. Since 1991, he has been a Professor of Electrical and Computer Engineering at the University of Delaware. His research interests include the electrical and optical properties of ternary and quaternary alloys of the group-IV semiconductors, and their device and circuit applications. He has investigated aluminum nitroxide for transit-gate applications.

Olivier Gauthier-Lafaye is working toward the Ph.D. degree in physics at the Institut d'Electronique Fondamentale, Université Paris-Sud, Orsay, France, since 1996.

His primary subject of research is the development of optically pumped mid-IR sources in semiconductors heterostructures.

Sébastien Sauvage graduated from the École Polytechnique in 1994, received the Quantum Physics DEA (Diplôme d́ Étude Approfondi) at the École Normale Supérieure, Paris, in 1995, and is working toward the Ph.D. degree in physics at the Université Paris-Sud, Orsay, France.

Since 1996, he has been working at the Institut d'Electronique Fondamentale at the University of Orsay where his research tackles the IR optical properties of low-dimensional systems in III–V semiconductors.

J.-L. Perrossier was born in Rives, France, on August 17, 1950. He received the Dipl. Ing. degree in physico-chemistry analysis from the Conservatoire National des Arts et Metiers.

Currently, he is a CNRS Research Engineer at the Institut d'Electronique Fondamentale, Université Paris-Sud, Orsay, France. He is engaged in electrical characterization on metal–semiconductor junctions and on photodiodes on IV–IV semiconductors at room and low temperatures.

Philippe Boucaud received the M.S. degree in solid state physics and the Ph.D. degree in optical nonlinearities in semiconductor quantum wells from the University of Paris XI, in 1987 and 1992, respectively.

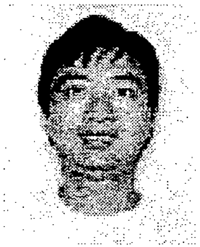
Since 1993, he has been a Permanent Researcher at the National Center of Scientific Research (CNRS). He is currently working at Institut d'Electronique Fondamentale which he has joined in 1988. He has been a Post-Doctoral Fellow at France Telecom CNET in 1992 where he has been working on the growth and optical properties of SiGeC heterostructures. His current research interest lies in the physics of intersubband transitions in quantum wells, the optical properties of self-organized quantum nanostructures and the structural and optical properties of silicon-based (SiGeC) quantum heterostructures. He has coauthored over 75 papers in the field of semiconductor heterostructures.

Francois H. Julien graduated from Ecole Centrale des Arts et Manufactures in 1978, received the masters diploma in 1980 and the Doctorat d'Etat (Ph.D.) in physics in 1985 from the Université Paris-Sud with research activities on optically pumped far-IR and mid-IR molecular lasers and on nonlinear optical interactions in gases.

He is Directeur de Recherche at the National Center for Scientific Research which he joined in 1981. From 1985 to 1987, he was a Visiting Scientist at University of Illinois at Urbana-Champaign with research on integrated optics in semiconductor superlattices and quantum wells. He started a group in 1987 at the Institut d'Electronique Fondamentale of the Université Paris-Sud on quantum optics in low-dimensional semiconductor structures with special emphasis on intersubband transitions and more recently unipolar lasers. His main interests include the optoelectronic properties of quantum wells, wires, and dots in III–V and IV–IV semiconductors. He is author or coauthor of more than 120 publications international journals.

Jean-Michel Lourtioz (SM'88) was born in Lens, France, in 1948. He graduated from Ecole Centrale de Paris in 1971, received the equivalent of the M.S. degree in physics and the Ph.D. degree from the University of Paris, in 1975 and 1981, respectively.

Since 1976, he has been with CNRS, and has worked at the Institut d'Electronique Fondamentale, University of Paris, France. He is presently at the head of a research Department and Directeur de Recherche at CNRS. His current research interests include semiconductor laser dynamics, nonlinear optics, intersubband transitions quantum wells and superlattices, group-IV heterostructures and devices. Since 1996, he is also in charge of research studies on photonic crystals and microcavities.



Fen Chen received the B.S. degree in physics in 1990 from Shanghai University of Science and Technology, the M.S. degree in physics in 1994 from North Dakota State University, Fargo, ND, and Ph.D. degree in electrical engineering in 1998 from University of Delaware, Newark. His dissertation was entitled "Electrical Characterization of SiGeC Alloy Materials and Devices."

In 1990, he worked as a process engineer at Shanghai Institute of Space Power-sources and engaged in high-efficiency Si solar cell research. He was a student intern in 1997 at Intel Corp. in Santa Clara, CA, and worked on layered metal films for multilevel interconnections. He is currently a Reliability Engineer at IBM Microelectronics Division in Essex Junction, VT.

Bradley A. Orner received the B.S. degree from Johns Hopkins University, Baltimore, MD, and the Ph.D. degree in electrical engineering from the University of Delaware, Newark.

His research focused on growth of SiGeC alloys by molecular beam epitaxy and optical charactarization of SiGeC alloys. He is currently with SiGe Technology Integration at IBM in Burlington, VT.

Kristofer Roe was born in Saratoga Springs, NY, in 1974. He received the B.E.E. degree in electrical engineering in 1995 and the M.E.E. degree in electrical engineering in 1997, respectively, from the University of Delaware, Newark, where he is working toward the Ph.D. degree

His current and past thesis work includes: X-ray diffraction studies of semiconductor films, polymer-semiconductor devices, and novel diode and HBT structures.

Cyril Guedj received the M.E. degree from the Institut des Sciences de la Matiere et du Rayonnement (ISMRa, ENSI-Caen) and the M.S. degree in material science from CAEN University, France. In 1992, he also obtained a M.B.A. degree from the Institut d'Administration des Entreprises (IAE, CAEN) and received the Ph.D. degree in microelectronics with highest honors from the University of Paris XI in 1997.

He won the E-MRS student award in 1995, His Ph.D. research focused on the growth, characterization, theoretical modeling and industrial applications of Si-Ge-C heterostructures on silicon. He also invented several devices and systems for Thomson, Atomic Energy Branch (CEA) and Alcatel Alsthom Recherche successively. He is currently a Post-Doctoral researcher at the University of Delaware in the department of Electrical and Computer Engineering, where he's studying experimental and theoretical properties of novel Si–Ge–C materials and devices.

R. G. Wilson studied engineering physics at Cornell University, Ithaca, NY and received the Ph.D. degree in nuclear physics from Ohio State University, Columbus, OH, in 1961.

He has 44 years experience in ion and electron beam technologies, especially ion implantation and secondary ion mass spectrometry, and especially applied to materials and devices for the semiconductor electronics and optics devices industry. He spent 35 of those years at Hughes Research Laboratories, 2 years at Rocketdyne in advanced space propulsion, and 4 years at Battelle Memorial Institute in nuclear reactor critical assembly development. Currently, he is a consultant. He has written 300 papers, five books, several NBS Special Publications, and received seven patents.

Jennifer Spear received the B.S. degree in materials science and engineering from the University of Pittsburgh, Pittsburgh, PA, in April 1993, and is working toward the M.S. degree in materials science and engineering with specialization in semiconductor processing and manufacturing at Arizona State University, Tempe.

She is an Associate Product Manager for the X-ray diffraction product line at Philips Analytical. At Philips, she works with a variety of thin films including, semiconductor, magnetic, and optical materials.

JOURNAL OF APPLIED PHYSICS VOLUME 85, NUMBER 1 1 JANUARY 1999

Strain modification in thin $Si_{1-x-y}Ge_xC_y$ alloys on (100) Si for formation of high density and uniformly sized quantum dots

Xiaoping Shao, Ralf Jonczyk, M. Dashiell, D. Hits, B. A. Orner, A.-S. Khan, K. Roe, J. Kolodzey, and Paul R. Berger[a]
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

M. Kaba and M. A. Barteau
Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716

K. M. Unruh
Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716

(Received 26 May 1998; accepted for publication 21 September 1998)

The effects of alloying C with Ge and Si and varying the C/Ge ratio during the growth of very thin layers of the ternary alloy SiGeC grown on Si (100) substrates and the resulting strain modification on self-assembled and self-organized quantum dots are examined. During coherent islanded growth, where dislocations are not formed yet to relieve the strain, higher strain energy produced by greater lattice mismatch acts to reduce the island size, increase the density of islands, and significantly narrow the distribution of island sizes to nearly uniformly sized quantum dots. Strain energy can also control the critical thickness for dislocation generation within the three-dimensional islands, which then limits the maximum height which coherent islands can achieve. After the islands relax by misfit dislocations, the island sizes increase and the island size distribution becomes broader with the increase of misfit and strain. The optimal growth for a high density of uniform coherent islands occurred for the $Si_{0.49}Ge_{0.48}C_{0.03}$ alloy composition grown on (100) Si, at a growth temperature of 600 °C, with an average thickness of 5 nm, resulting in a narrow size distribution (about 42 nm diameter) and high density (about 2×10^{10} dots/cm^2) of quantum dots. © *1999 American Institute of Physics.* [S0021-8979(99)00801-4]

I. INTRODUCTION

Since the advent of molecular beam epitaxy over two decades ago, the growth of semiconductors with monolayer control has attracted enormous attention for quantum confined semiconductor structures.[1–3] Low dimensional structures have generated tremendous interest at many levels including the study of their fundamental physics as well as potentially important technological applications in electronic and optoelectronic devices due to their new electronic and optical properties. Currently there has been a tremendous commercial success in applications of quantum well based devices.[4] Quantum wires and quantum dots have numerous advantages over quantum wells. For example, quantum dots allow single charge counting and tunneling,[5–7] and the higher confinement modifies the density-of-states to significantly lower laser threshold currents and elevate differential gain. The fabrication of quantum wires and quantum dots with high densities and high uniformity, however, is difficult. Some experimental methods require complex and expensive lithography and processing. An alternative approach is self-organization by directly controlling the two-dimensional (2D) to 3D growth transition induced by lattice misfit stress. Studies on the surface morphology of heteroepitaxial films showed that the growth kinetics are strongly influenced by lattice mismatch induced strain energy, surface free energy, growth temperature, and growth rate.[8–10]

SiGe alloys seem to be a promising material system for realizing quantum dot structures for terabyte storage applications, since they can readily be implemented within existing Si technology. The growth of Ge on Si follows the Stranski–Krastanov mode of self-organized Ge-island formation when the epilayer thickness exceeds a few monolayers.[11–16] The addition of small amounts of C to SiGe acts to compensate in approximately a 1:8 ratio the compressive strain created by the addition of Ge to Si. This study examines alloying C with SiGe to form novel SiGeC quantum dots; the effects of subsequent strain modification on self-assembled and self-organized quantum dots are examined with improved control over strain.

II. EXPERIMENTAL PROCEDURES

A series of four $Si_{1-x-y}Ge_xC_y$ islanded samples were sequentially grown by the molecular beam epitaxy (MBE) technique in an EPI 620 solid source MBE system.[17,18] Silicon (100) wafers were prepared by chemically oxidizing the silicon surface using a solution of $H_2O:H_2O_2:HCl$ (7:5:5), followed by dipping the wafers in a dilute hydrofluoric acid solution (10:1 $H_2O:HF$). After the HF dip, the samples were immediately loaded into the MBE load lock chamber to be pumped down to 1×10^{-8} Torr. The wafers were then trans-

---

[a] Electronic mail: pberger@ee.udel.edu

© 1999 American Institute of Physics

Appendix 40

TABLE I. Summary of growth parameters, calculated critical thicknesses and measured TEM and AFM parameters for the four $Si_{1-x-y}Ge_xC_y$ epilayers grown on (100) Si used in this study.

Composition	$Ge_{0.97}C_{0.03}$	$Si_{0.39}Ge_{0.58}C_{0.03}$	$Si_{0.49}Ge_{0.48}C_{0.03}$	$Si_{0.7}Ge_{0.27}C_{0.03}$
Sample ID	SGC224	SGC225	SGC226	SGC227
Growth time (min)	55	55	65	95
Epilayer thickness (Å)	50	50	50	50
Critical thickness (Å)	22	67	107	1735
Misfit (%)	4.0	1.4	1.0	0.1
No. of grains in 4×4 μm^2	77	2552	3171	1839
Grain size mean (nm^2)	2.5×10^4	1.9×10^3	1.4×10^3	4.3×10^3
Grain size diameter (nm)	178	49	42	74
Wavelength (μm/cycle)	0.5	0.078	0.068	0.085
Mean roughness (nm)	9.6	4.8	3.0	4.8
Peak relaxation misfit (%)	2.0	0.7	0.5	0.05
Modified crit. thickness in 3D island (Å)	40	162	248	3800
Island type	Relaxed	Relaxed	Coherent	Coherent

ferred into the growth chamber where they were prebaked at 200 °C following a procedure similar to that discussed in Eaglesham *et al.*[19] The prebake was shown to effectively desorb hydrocarbons on the Si surface which resulted in high quality Si epitaxy for substrate temperatures of ⩾370 °C (defect densities $<10^5$ cm^{-2}). This step is necessary to avoid carbon contamination at the surface, which may act as nucleation sites for 3D growth. The samples were ramped to the growth temperature and the (2×1) reconstructed Si surface was confirmed by reflection high energy electron diffraction (RHEED). The substrate temperature during growth was held at 600 °C. The nominal film thickness for all the layers, assuming 2D growth, was 5 nm. Growth of the thin layers was performed without a buffer layer, but after RHEED confirmation of the (2×1) stabilized Si surface. The growth conditions maintained a nominal carbon composition of 3% for all of the four samples, while varying the Ge:Si ratio. The Si content was varied among 0%, 39%, 49%, and 70% (Table I), while the nominal C composition was held at 3%.

III. RESULTS AND DISCUSSION

The alloy composition of these epilayers was not measured directly, due to the very thin epitaxial layers employed here and the uneven step coverage. The alloy compositions were estimated from the growth conditions used, which were calibrated from Rutherford backscattering (RBS) measurements of thicker samples grown under similar conditions. A suitable method for measuring substitutional C in SiGeC alloys, without assuming Vegard's law holds, is to perform C resonant RBS. However, resonant RBS cannot measure alloy compositions below 1% C or thicknesses below ~1000 Å, and its error is ±0.2% above this threshold. Previous absorption measurements of $Si_{1-x}C_x$ and $Si_{1-x-y}Ge_xC_y$ layers monitoring the 810–815 cm^{-1} line associated with the vibration of coherent SiC precipitates and the 602–610 cm^{-1} line associated with the vibrational mode of substitutional C in Si have shown significant substitutional C for growth temperatures employed in this study.[20,21] Note the spread in the local vibrational mode energies due to the alloy effect with Ge.

The misfit dislocations at the substrate interface of these islands were examined with a Philips 400T transmission electron microscope (TEM), operating at 100 kV. Plan-view images were taken at the (220) Bragg reflection. Moiré fringes parallel to the (220) reflection plane were only found in the $Ge_{0.97}C_{0.03}$ sample (SGC224) and relatively large islands (diameter ⩾50 nm) of the $Si_{0.39}Ge_{0.58}C_{0.03}$ sample (SGC225), indicating these islands are relaxed by interfacial misfit dislocations. No dislocations (or Moiré fringes) were observed in any islands for the $Si_{0.49}Ge_{0.48}C_{0.03}$ (SGC226) and $Si_{0.7}Ge_{0.27}C_{0.03}$ (SGC227) epilayers, and a strain contrast can be seen in both images, indicating that the islands in these two samples are coherently strained.

Atomic force microscopy (AFM) was used in the tapping mode to examine the surface morphology of each sample, and catalog the island size and distribution. Figure 1 shows typical AFM topographies from all four sample compositions: $Ge_{0.97}C_{0.03}$ (SGC224), $Si_{0.39}Ge_{0.58}C_{0.03}$ (SGC225), $Si_{0.49}Ge_{0.48}C_{0.03}$ (SGC226), and $Si_{0.7}Ge_{0.27}C_{0.03}$ (SGC227). The 3D island size distribution of these samples is also shown in histogram form (Fig. 1). Note that as the overall Si content increases, the misfit and strain energy is reduced, and AFM analysis shows that this acts to reduce the island size, increase the island density, and tighten the variation in island size, for the two relaxed islanded samples: SGC224 and SGC225. However, sample SGC226, with coherent islands, has the highest density of islands (about 2×10^{10} dots/cm^2), smallest island size (about 42 nm diameter), and the most uniform size distribution in this series. The sample with the largest Si content, $Si_{0.7}Ge_{0.27}C_{0.03}$ (SGC227) which also has coherent islands, shows an increased island size, a decreased island density, and a broader island size distribution, compared to SGC226. This trend is also observed in the power spectrum density analysis of the AFM data, shown in Fig. 2, which is the Fourier transform of the measured raw data, where the peaks shift to a smaller wavelength with increasing Si composition, but shifts back to larger wavelengths for the $Si_{0.7}Ge_{0.27}C_{0.03}$ sample. The changes in surface roughness with Si content can also be

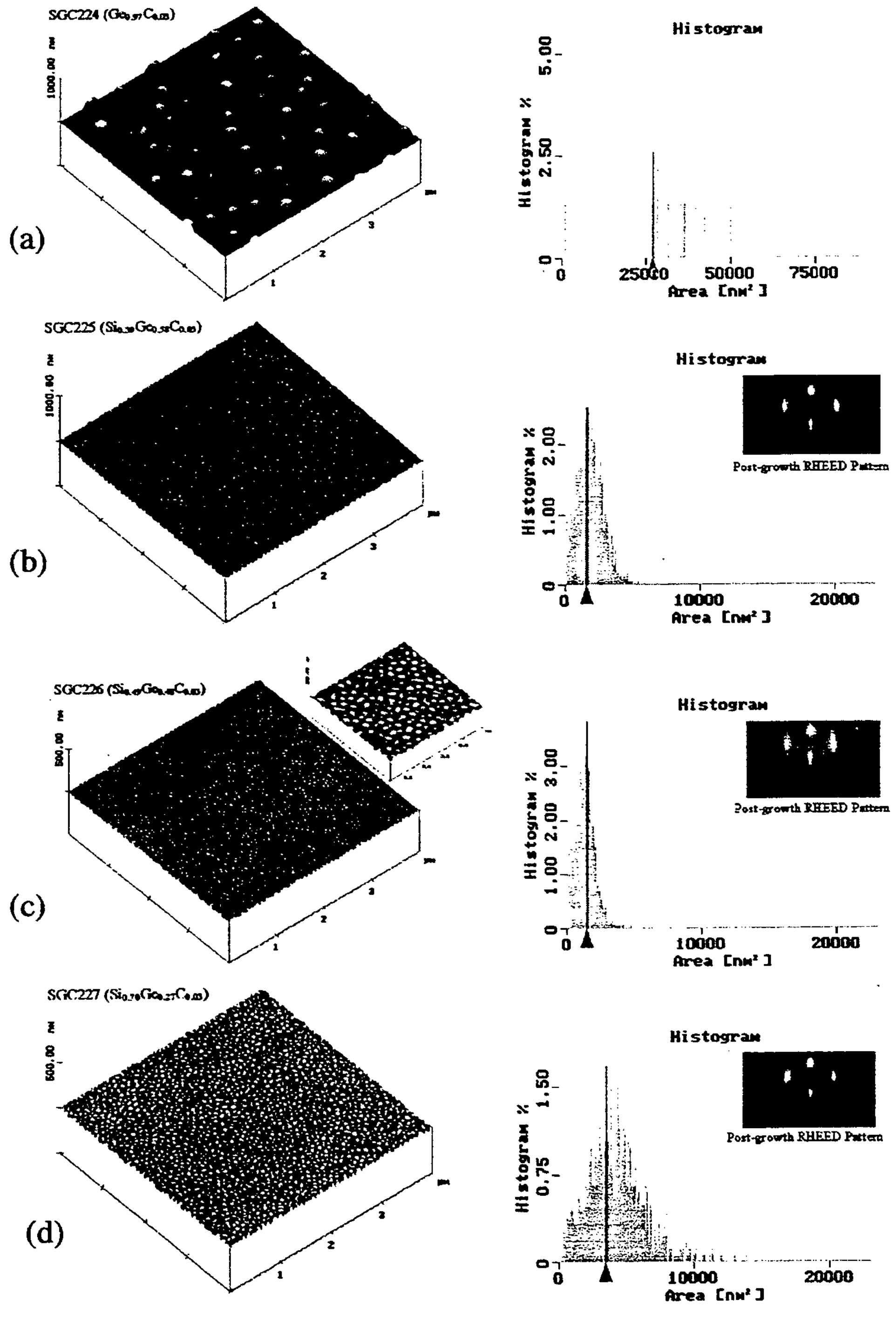


FIG. 1. AFM topographs of: (a) $Ge_{0.97}C_{0.03}$ (SGC224), (b) $Si_{0.39}Ge_{0.58}C_{0.03}$ (SGC225), (c) $Si_{0.49}Ge_{0.48}C_{0.03}$ (SGC226), and (d) $Si_{0.7}Ge_{0.27}C_{0.03}$ (SGC227), and their corresponding histograms, showing island size distribution. Insets show their corresponding postgrowth RHEED pattern.

seen from *in situ* reflection high-energy electron diffraction (RHEED), as shown in the insets in Figure 1.

Low temperature photoluminescence (PL) studies of all four layers are shown in Fig. 3. In all the layers a peak at 1.09 eV was observed which is related to the Si substrate. Narrow peaks are observed at 0.789 and 0.767 eV, as well as a broad band centered at 0.75 eV in the layers that contain silicon. They were found to increase in intensity with increasing Si content and decreasing Ge/C ratio. Very similar luminescence was observed by Wang *et al.*,[22] and was attributed to Ge no-phonon lines. However, studies of thermally induced defects in silicon by Minaev *et al.*[23] and Weber *et al.*[24] revealed luminescence at the same locations. They observed that electron–vibrational emission bands, *P* (0.767 eV), *H* (0.9258 eV), *T* (0.9356 eV) and *I* (0.9653 eV), found in *n*-type and *p*-type silicon caused thermally induced de-

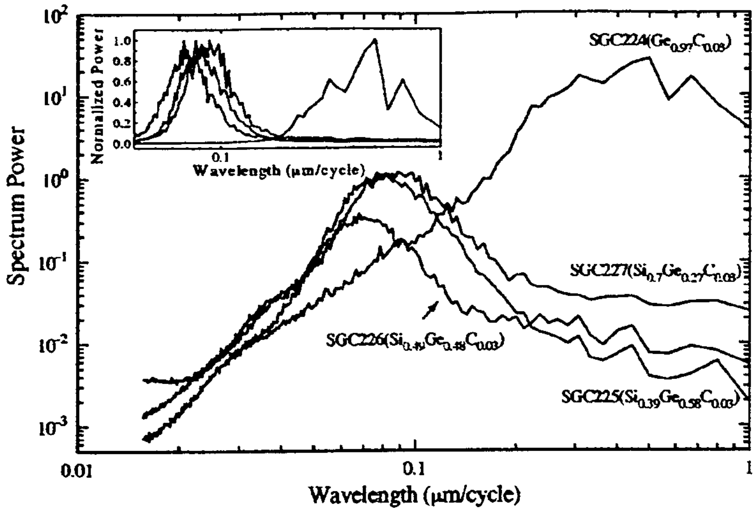


FIG. 2. Power spectrum density, showing Fourier spectral intensities of: $Ge_{0.97}C_{0.03}$ (SGC224), $Si_{0.39}Ge_{0.58}C_{0.03}$ (SGC225), $Si_{0.49}Ge_{0.48}C_{0.03}$ (SGC226), and $Si_{0.7}Ge_{0.27}C_{0.03}$ (SGC227), grown on (100) Si.

fects, incorporating oxygen and carbon impurity atoms. Additionally, they found the *P* line increased in intensity when the oxygen concentration was substantially larger than the carbon concentration. We attribute these PL lines in the SiGeC layers to thermally induced defects. The reason may be due to the thermal limitations on the Si effusion cell used in this study. Due to the ceramic crucible used for the Si source, the Si furnace cannot be raised too high in temperature. As a result, the epitaxial growth rate drops off considerably with increasing Si content. As the growth rate is reduced, the likelihood of impurity incorporation increases. For this reason, the oxygen contamination climbs with increasing Si content, and probably deleteriously affects the growth front of the last sample (SGC227), which is expected to exhibit a layer-by-layer 2D growth front with only a 0.1% misfit.

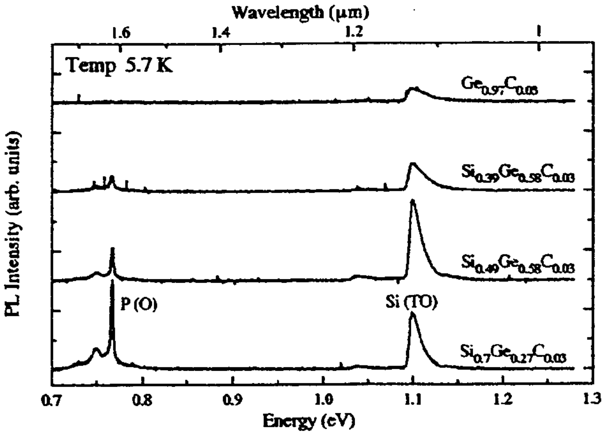


FIG. 3. Photoluminescence (PL) spectrum of: $Ge_{0.97}C_{0.03}$ (SGC224), $Si_{0.39}Ge_{0.58}C_{0.03}$ (SGC225), $Si_{0.49}Ge_{0.48}C_{0.03}$ (SGC226), and $Si_{0.7}Ge_{0.27}C_{0.03}$ (SGC227), grown on (100) Si. Note, the elevated *P* line due to impurity incorporation, notably, oxygen, as the growth rate is reduced for the higher concentrations of Si.

IV. CONCLUSIONS

Our results show that during coherent islanded growth, a larger Ge content, and therefore greater strain, can reduce the island size, increase island density, and narrow the island size distribution. Strain can also control the critical thickness for dislocation generation within 3D islands, which limits the maximum height that coherent islands can achieve. After relaxation by misfit dislocations, the island size increases and the island size distribution becomes broader with the decrease of Si content and increase of strain. In our case, the optimal growth for a high density of uniform coherent islands occurred for the composition of $Si_{0.49}Ge_{0.48}C_{0.03}$ on (100) Si, at a growth temperature of 600 °C, for an average thickness of 5 nm, resulting in a narrow size distribution (about 42 nm diameter) and high density (about 2×10^{10} dots/cm^2).

ACKNOWLEDGMENTS

The authors wish to express their thanks to Dr. Robert Hull of the University of Virginia for his fruitful discussions. This work was supported by the National Science Foundation (Grant No. ECS-9624160).

[1] A. Y. Cho and J. R. Arthur, Prog. Solid State Chem. **10**, 157 (1975).
[2] L. Esaki and L. L. Chang, Phys. Rev. Lett. **33**, 495 (1974).
[3] R. Dingle, W. Wiegmann, and C. H. Henry, Phys. Rev. Lett. **33**, 827 (1974).
[4] C. Weisbuch and B. Vinter, *Quantum Semiconductor Structures* (Academic, New York, 1991).
[5] D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. Frost, D. G. Hasko, D. C. Peacock, and C. T. Foxon, J. Phys. C **21**, L209 (1988).
[6] A. Imamoglu and Y. Yamamoto, Phys. Rev. Lett. **72**, 210 (1994).
[7] *Single Charge Tunneling*, NATO Advanced Science Instrument Series B, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1992).
[8] W. A. Jesser and J. H. van der Merwe, Surf. Sci. **31**, 229 (1972).
[9] P. R. Berger, K. Chang, P. Bhattacharya, J. Singh, and K. K. Bajaj, Appl. Phys. Lett. **53**, 684 (1988).
[10] J. Pamulapati, P. K. Bhattacharya, J. Singh, P. R. Berger, C. W. Snyder, B. G. Orr, and R. L. Tober, J. Electron. Mater. **25**, 479 (1996).

[11] H. Sunamura, N. Usami, Y. Shiraki, and S. Fukatsu, Appl. Phys. Lett. **66**, 3024 (1995).

[12] P. Schittenhelm, M. Gail, J. Brunner, J. F. Nützel, and G. Abstreiter, Appl. Phys. Lett. **67**, 1292 (1995).

[13] E. Palange, G. Capellini, L. DiGaspare, and F. Evangelisti, Appl. Phys. Lett. **68**, 2982 (1996).

[14] M. Krishnamurthy, B.-K. Yang, and W. H. Weber, Appl. Phys. Lett. **69**, 2572 (1996).

[15] O. G. Schmidt, C. Lange, K. Eberl, O. Kienzle, and F. Ernst, Appl. Phys. Lett. **71**, 2340 (1997).

[16] R. Jonczyk, D. A. Hits, L. V. Kulik, J. Kolodzey, M. Kaba, and M. A. Barteau, J. Vac. Sci. Technol. B **16**, 1142 (1998).

[17] J. Kolodzey *et al.*, J. Cryst. Growth **157**, 386 (1995).

[18] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, Appl. Phys. Lett. **67**, 1865 (1995).

[19] D. J. Eaglesham, G. S. Higashi, and M. Cerullo, Appl. Phys. Lett. **59**, 685 (1991).

[20] M. W. Dashiell, L. V. Kulik, D. Hits, J. Kolodzey, and G. Watson, Appl. Phys. Lett. **72**, 833 (1998).

[21] L. V. Kulik, D. A. Hits, M. W. Dashiell, and J. Kolodzey, Appl. Phys. Lett. **72**, 1972 (1998).

[22] X. Wang, Z. Jiang, H. Zhu, F. Lu, D. Huang, X. Liu, C. Hu, Y. Chen, Z. Zhu, and T. Yao, Appl. Phys. Lett. **71**, 3543 (1997).

[23] N. S. Minaev and A. V. Mudryi, Phys. Status Solidi A **68**, 561 (1981).

[24] J. Weber and R. Sauer, Mater. Res. Soc. Symp. Proc. **14**, 165 (1983).

APPLIED PHYSICS LETTERS VOLUME 74, NUMBER 5 1 FEBRUARY 1999

Substitutional Ge in 3C–SiC

C. Guedj[a] and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

(Received 17 August 1998; accepted for publication 23 November 1998)

The incorporation of substitutional Ge into 3C–SiC alloys is studied theoretically with an anharmonic Keating model specifically adapted to the computation of the structural properties and the lattice dynamics of $Si_{1-x-y}Ge_xC_y$ alloys. Basic energy calculations show that the substitution of Si by Ge is more probable than the substitution of C by Ge in the zinc-blende silicon carbide crystal. If Ge replaces only Si, then the lattice parameter equals $(0.43593\pm0.00002)+(0.000337\pm0.000002)y$, where y stands for the Ge content. Hence, Vegard's law is not applicable. The alloy is characterized by a distinct phonon spectrum whose maximum peak position in cm^{-1} is best described by the exponential decay $(243\pm1)+(27\pm2)\exp[-y/(7.5\pm1.2)]$ up to the zinc-blende GeC compound. © *1999 American Institute of Physics.* [S0003-6951(99)01305-4]

Silicon carbide (SiC)-based electronics are suitable for high-power and high-frequency applications, as well as for severe environments, including high temperature and high radiation.[1,2] Silicon carbide's ability to function under such extreme conditions is expected to enable significant improvements in a far ranging variety of applications and systems. Substitutional Ge may modify the structural, electrical, and optical properties of SiC-based heterostructures, therefore, Ge incorporation into SiC-based microelectronics and optoelectronics may provide[3] further device opportunities through band-gap and strain engineering. The crystal growth might not be simple, mainly because of the difference in lattice parameters and covalent radii between silicon germanium and diamond, but Ge seems to have a beneficial effect on the epitaxy of single-crystalline 3C–SiC on silicon.[4] There are few if any data available about substitutional Ge in 3C–SiC to our knowledge.[4] Because of the broad technological importance of IV–IV materials and devices, as well as the increasing ability to grow highly metastable alloys, it is essential to develop theoretical predictions of the physical properties of this material.

The understanding of the lattice-vibrational properties is important to explain various interesting properties of SiC, among them mechanical, thermal, and structural ones. For example, phonons may stabilize the polytypism via several contributions to the free energy.[5] Here, the structural properties and lattice dynamics of substitutional Ge in 3C–SiC are determined using a modified anharmonic Keating model specially adapted to the computation of lattice parameters and phonon spectra of $Si_{1-x-y}Ge_xC_y$ alloys. First, the theoretical model will be briefly described, then the structural effects of substitutional Ge in 3C–SiC and the corresponding localized vibrational spectra will be successively detailed.

Our theoretical approach is based upon a valence force-field model derived from Keating[6] and taking into account the effects of carbon anharmonicity. The interactions between atoms have been modeled with an interatomic potential similar to the one of Rücker *et al.*[7] with the exception of adjusting our force coefficients to yield the correct lattice parameters and phonon modes of Si, Ge, $Si_{1-x}Ge_x$, diamond, and 3C–SiC at 300 K. Our model gives a localized vibrational mode of C in pure Ge situated at 531 cm^{-1}, in agreement with Hoffmann *et al.*[8] The very high precision and reliability of these experimental data, which can be obtained by x-ray diffractometry, Raman spectrometry, and absorption spectroscopy, justify our approach. In addition, our set of parameters enables the computation of the LO(Γ) mode from 3C–SiC, which cannot be obtained from the parameters given by Rücker *et al.*,[7] to our knowledge. This model is, therefore, ideally suited to a precise computation of lattice parameters and phonon spectra of tetrahedrally coordinated $Si_{1-x-y}Ge_xC_y$ alloys. The molecular dynamics relaxation is calculated from a 512 atom supercell by time increments of 3 fs, until the total potential energy reaches a stable minimum. We have computed isotropic relaxations, where no external pressure is applied to the computational box, to simulate the structural properties of the alloy. In a further stage, the local phonon density around carbon is computed, using the recursion method detailed in Ref. 9. This algorithm enables the computation of the local phonon density around chosen atoms. The calculated spectra can be compared with absorption spectroscopy and Raman spectrometry results after application of proper selection rules. Numerous atomic configurations have been tested, in order to simulate the substitution of Ge into the 3C–SiC crystal.

The application of these theoretical tools gives an insight into the structure of the substitutional alloy. In the case of a single substitutional Ge into the zinc-blende SiC matrix, the atomic positions after isotropic relaxation are depicted in Fig. 1. The substitution of C by Ge induces a higher lattice distortion than the substitution of Si, which is logical because Ge is close to Si in terms of atomic radius, electronegativity, and elastic properties. The substitution of C requires at least 38 times more energy, therefore, it is highly unprobable. Silicon carbide is more rigid than silicon, therefore, the distortions are confined across few atomic distances. This strain modifies the crystal dimensions, in a manner depending on the statistical atomic distribution. The lattice param-

---

[a] Corresponding author. Electronic mail: guedj@ief.u-psud.fr

© 1999 American Institute of Physics

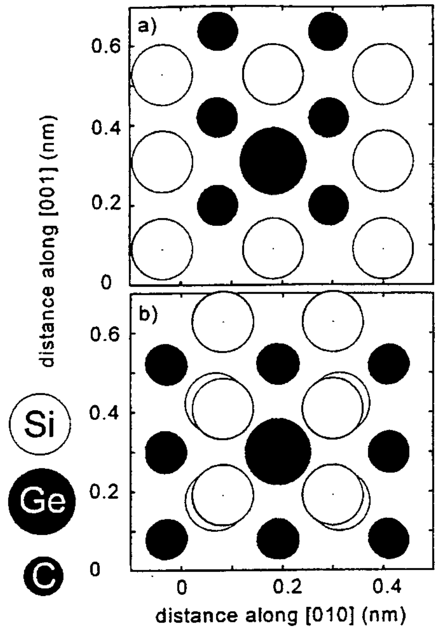


FIG. 1. View along [100] of a substitutional Ge into 3C–SiC. Molecular dynamics computation is used to relax the 512 atoms supercell isotropically. C, Si, and Ge are represented with spots of increasing diameter. (a) is obtained when Ge replaces Si, while (b) corresponds to the substitution of C by Ge. A modified Keating model which takes into account the anharmonicity of C is used to compute the atomic interactions.

eter is obviously proportional to the mean size of the supercell after relaxation, and the results are displayed in Fig. 2. If we assume that Ge replaces Si only [Fig. 2(a)], then the lattice parameter equals $(0.43593\pm0.00002)+(0.000337\pm0.000002)y$, where y stands for the Ge content. A Vegard's law [Fig. 2(b)] between Ge and cubic SiC would give a lattice parameter of $0.43596+0.0012994x$. Even in the hypothesis of a random substitution of Si or C by Ge [Fig. 2(c)], Vegard's law does not apply. In the case of a random substitution, the increase of the lattice parameter is rapid, but the corresponding increase of elastic energy is unlikely to be observed experimentally, because more stable atomic arrangements will occur instead. In any case, the modification of the lattice parameter induced by substitutional Ge in 3C–SiC should be easily probed by x-ray diffraction.

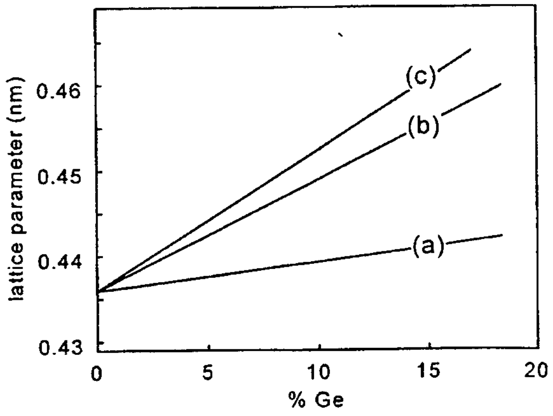


FIG. 2. Evolution of the lattice parameter after incorporation of substitutional Ge in 3C–SiC. (a) is obtained when Ge replaces only Si in a random manner. (b) is computed using a linear combination of the lattice parameters of Si, Ge, and diamond, considering the relative concentrations (Vegard's law). (c) is obtained when Ge replaces Si or C randomly.

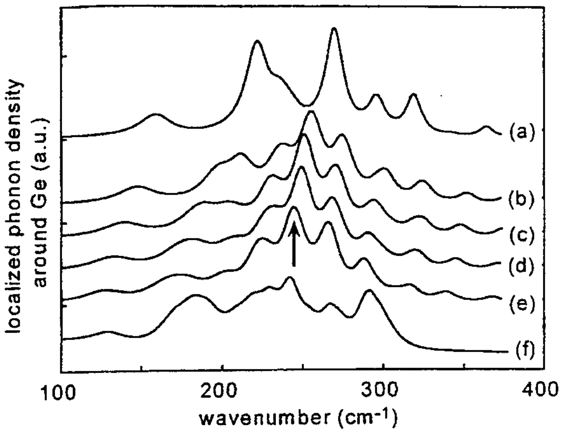


FIG. 3. Localized phonon spectra around Ge calculated by the recursion method. Ge is assumed to replace only Si in the zinc-blende 3C–SiC crystal. Curves (a)–(f) are calculated for increasing substitutional Ge concentrations (respectively, 0.2%, 4.9%, 9.4%, 14.8%, 18.4%, and 50%).

The computation of phonon spectra give a further tool to investigate this material. In Figs. 3(a)–3(f), the localized vibrational spectra around Ge are computed for increasing Ge content. The substitutional incorporation of Ge in 3C–SiC is characterized by a distinct phonon spectrum whose maximum peak position in cm^{-1} is best described by the exponential decay $(243\pm1)+(27\pm2)\exp[-y/(7.5\pm1.2)]$ up to the zinc-blende GeC compound. Because of the strong local elastic heterogeneity, the calculated spectrum is usually complex, and the clear maximum obtained in curve (e) of Fig. 3 is marked by an arrow. In the real crystal, a smoothing of the spectrum should naturally occur because of the very high number of atoms involved. Hence, our fit assumes an additional Gaussian smoothing. Physically, these spectra could be considered as a result of complex multiphonon processes involving the nearest neighbors of Ge. The Ge–C bond is stabilized in the zinc-blende GeC compound, therefore, it is reasonable that the fitted phonon energy decreases when the Ge concentration increases. According to the lattice dynamics calculations, this peak should be Raman active. Hence, proper substitutional Ge incorporation into 3C–SiC could be assessed by Raman spectrometry.

In summary, we have performed molecular dynamics simulations using a valence force-field model to simulate the incorporation of Ge into substitutional sites of 3C–SiC. The silicon site is energetically favored over its C counterpart for the Ge substitution, and the lattice parameter is expected to deviate from Vegard's law. If Ge replaces only Si, then the calculated lattice parameter equals $(0.43593\pm0.00002)+(0.000337\pm0.000002)y$, where y stands for the Ge content. The alloy is characterized by a distinct phonon spectrum whose maximum peak position in cm^{-1} is best described by the exponential decay $(243\pm1)+(27\pm2)\times\exp[-y/(7.5\pm1.2)]$ up to the zinc-blende GeC compound. This mode should be Raman active. Ge-doped silicon carbide, which can be probed by x-ray diffraction and Raman spectrometry, might play an important role in future IV–IV microelectronics and optoelectronics.

The authors gratefully acknowledge support from ARO under Grant No. DAAH04-95-1-0625, and DARPA Contract No. F49620-96-C-0006.

[1] R. F. Davis, Z. Sitar, B. E. Williams, H. S. Kong, H. J. Kim, J. W. Palmour, J. A. Edmond, J. Ryu, J. T. Glass, and C. H. Carter, Mater. Sci. Eng. B **B1**, 77 (1988).

[2] W. J. Choyke, in *The Physics and Chemistry of Carbides, Nitrides and Borides, Nato Advanced Study Institute, Series E: Appl. Sci. Vol. 185,* edited by R. Freer (Kluwer, Dordrecht, 1990).

[3] G. Katulka, C. Guedj, J. Kolodzey, M. W. Tsao, J. Rabolt, R. G. Wilson, and C. P. Swann, Appl. Phys. Lett. (in press).

[4] T. Hatayama, N. Tanaka, T. Fuyuki, and H. Matsunami, Appl. Phys. Lett. **70**, (1997).

[5] B. Winkler, M. T. Dove, E. K. H. Salje, M. Leslie, and B. Palosz, J. Phys.: Condens. Matter **3**, 539 (1991).

[6] P. N. Keating, Phys. Rev. **145**, 637 (1966).

[7] H. Rücker, M. Methfessel, B. Dietrich, K. Pressel, and H. J. Osten, Phys. Rev. B **53**, 1302 (1996).

[8] L. Hoffmann, J. C. Bach, B. Bech Nielsen, P. Leary, R. Jones, and S. Öberg, Phys. Rev. B **55**, 1167 (1997).

[9] A. Hairie, F. Hairie, G. Nouet, E. Paumier, and A. P. Sutton, in *Polycrystalline Semiconductors III—Physics and Technology*, edited by H. P. Strunk, J. H. Werner, B. Fortin, and O. Bonnaud, Vol. 37-38, 91 (1993).

APPLIED PHYSICS LETTERS VOLUME 74, NUMBER 8 22 FEBRUARY 1999

Growth and thermal stability of pseudomorphic $Ge_{1-y}C_y$/Ge superlattices on Ge(001)

R. Duschl,[a] O. G. Schmidt, W. Winter, and K. Eberl
Max-Planck-Institut für Festkörperforschung, Heisenbergstraße 1, 70569 Stuttgart, Germany

M. W. Dashiell and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

N. Y. Jin-Phillipp and F. Phillipp
Max-Planck-Institut für Metallforschung, Heisenbergstraße 1, 70569, Stuttgart, Germany

(Received 13 October 1998; accepted for publication 22 December 1998)

High quality Ge/$Ge_{1-y}C_y$ superlattices with nominal carbon contents of 1.2% and 2.1% were grown by molecular beam epitaxy on Ge(001). In transmission electron microscopy the layers are planar and perfectly pseudomorphic without any extended defects observable. The infrared absorption line at 529 cm^{-1} is attributed to the local vibrational mode of substitutional carbon in germanium. However, in contrast to $Si_{1-y}C_y$ alloys where almost 100% of the C is substitutional under optimized growth conditions, x-ray diffraction measurements indicate that the efficiency of carbon incorporation onto substitutional sites is only about 30% for low temperature growth at T_S =200 °C. It reduces further for higher growth temperatures to only about 10% at T_S=300 °C. Post-growth annealing experiments indicate thermal stability up to 450 °C. Annealing at higher temperature results in a reduction of substitutional carbon content. As in the case of $Si_{1-y}C_y$ alloys the built-in strain is relaxed by C diffusion and not by nucleation of misfit dislocation. © *1999 American Institute of Physics.* [S0003-6951(99)02108-7]

The substitutional incorporation of small amounts of carbon into Si and $Si_{1-x}Ge_x$ is an attractive method to adjust strain, band alignments and energy gaps in Si/SiGeC heterostructures.[1–3] For Ge-rich $Si_{1-x-y}Ge_xC_y$ layers, the incorporation of carbon into the Ge matrix plays an important role. Recent experiments on $Si_{1-x-y}Ge_xC_y$ films indicate that growth becomes increasingly more difficult for high Ge contents, implying that C tries to avoid C–Ge bonds.[4] Very careful optimization of the growth parameters in molecular beam epitaxy (MBE) is essential for the synthesis of $Ge_{1-y}C_y$ alloy layers, because of the even lower solubility of carbon in Ge (10^8 atoms/cm^3 at the melting point of Ge) as compared to Si, and due to the large lattice mismatch between diamond and Ge of about 59%. Previous investigations have focused on the synthesis and the material parameters of $Ge_{1-y}C_y$ layers on Si(001). Since pseudomorphic $Ge_{1-y}C_y$ alloys with only a few atomic layers can be prepared on Si,[5] most of the measurements were carried out on thick relaxed layers.[6–8] These layers exhibit high dislocation densities which makes it difficult to measure intrinsic material parameters.

In this letter, we report on the epitaxial growth of high quality $Ge_{1-y}C_y$/Ge superlattices (SL) with carbon concentrations up to 2.1% on Ge(001) substrates. The layers are well defined and allow detailed investigation of structural properties, efficiency of the incorporation of substitutional carbon and the thermal stability of the metastable $Ge_{1-y}C_y$ alloy.

The $Ge_{1-y}C_y$/Ge SL structures were grown by solid source MBE. The Ge(001) substrates were chemically cleaned in boiling trichlorethylene, acetone, and methanol and then oxidized in a H_2O_2/H_2SO_4 mixture. The oxide was desorbed in the MBE chamber by annealing at 550 °C for 10 min. The growth was started at a temperature of 150 °C to avoid the formation of islands. During a 50 nm thick buffer layer the temperature was increased up to 200 °C. At this temperature the SL structure, consisting of 30 periods alternating 10 nm Ge and 3 nm $Ge_{1-y}C_y$, and finally, a 20 nm Ge cap layer were deposited. Two structures, called A and B, were fabricated with nominal carbon contents of 1.2% and 2.1%, respectively. The growth rates were previously calibrated by x-ray diffraction (XRD) measurements of a $Si/Si_{1-x}Ge_x$ and two $Si/Si_{1-y}C_y$ SL. The $Si/Si_{1-y}C_y$ samples were grown at a growth rate of 1 Å/s and a substrate temperature of 460 °C. The carbon content, calculated with the approach by Kelires,[9] was 0.45% and 0.80%, respectively. The flux calibration by XRD is reasonable because former measurements with secondary-ion mass spectroscopy and XRD showed that nearly 100% of the carbon atoms were incorporated at substitutional sites for these growth conditions, which is in agreement with recent investigations by Zerlauth *et al.*[10] For the deposition of germanium, we used an effusion cell and carbon was sublimated from a hot pyrolytic graphite filament. Transmission electron microscopy (TEM), double-crystal XRD, using Cu $K\alpha_1$ radiation, and infrared (IR) absorption measurements were applied to analyze the structural properties and the amount of substitutional C. To investigate the thermal stability the samples were annealed in a rapid thermal annealer (RTA) in forming gas atmosphere in the temperature range from 300 to 850 °C for various times. X-ray rocking curve analysis of the symmetric (004) and the asymmetric (115) reflections was used to cal-

---

[a] Electronic mail: duschl@servix.mpi-stuttgart.mpg.de

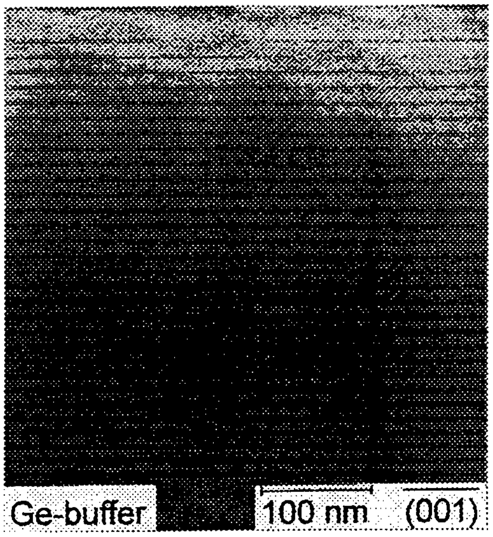


FIG. 1. Cross-sectional TEM image of 30 periods alternating 10 nm Ge and 3 nm $Ge_{0.979}C_{0.021}$. The image indicates no signs of extended C related defects or misfit dislocations.

culate the amount of substitutional carbon and the degree of relaxation of the layers.[11]

Figure 1 shows a cross-sectional TEM image of the sample B with 3 nm thick $Ge_{0.979}C_{0.021}$ layers. The $Ge_{0.979}C_{0.021}$ layers appear as homogeneous dark lines with planar interfaces. The image does not indicate any extended defects induced by the carbon. Figure 2(a) shows XRD curves of the two $Ge_{1-y}C_y$/Ge SL A and B described above. The main peak at $\Delta 2\Theta = 0°$ is the (004) diffraction from the Ge substrate and buffer layer. The zero-order peak of the SL is shifted to values of $\Delta 2\Theta = +0.058°$ and $\Delta 2\Theta = +0.084°$ for samples A and B, respectively, indicating that the average strain is tensile. Evidence of good crystalline quality is reflected in the observation of higher order SL peaks and observable Pendelosung fringes around the zero-order peak [see inset of Fig. 2(a) for sample B]. The TEM image in Fig. 1 and the evaluation of the asymmetric (115) rocking curves demonstrate that the $Ge_{1-y}C_y$ alloy layers are fully laterally extended to match the Ge substrate (pseudomorphic growth). Figure 2(a) shows the best fitted simulated curves. The dynamical simulation uses Vegard's law, i.e, the linear interpolation between Ge and diamond, to calculate the lattice constant $a_0(y)$ of $Ge_{1-y}C_y$. A slight bowing effect in $a_0(y)$, similar to the case of $Si_{1-y}C_y$,[9] would reduce the carbon concentration indicated in the following, but does not change the other statements. The results of the fitting procedures for the samples A and B are carbon contents of 0.5% and 0.8%, respectively. We have compared this C content with the nominal carbon flux calibrated by XRD measurements on $Si_{1-y}C_y$ SL structures. A comparison with other methods, like secondary-ion mass spectroscopy measurements, is difficult because of missing standards for $Ge_{1-y}C_y$. The comparison of the C flux with the C content of the $Ge_{1-y}C_y$ layers results in an efficiency of about 30% for the incorporation of substitutional C into Ge at $T_S = 200$ °C. For higher growth temperatures of 300 °C, the substitutional C incorporation is reduced to only 10%. Significantly lower T_S, on the other hand, will result in increased incorporation of point defects.

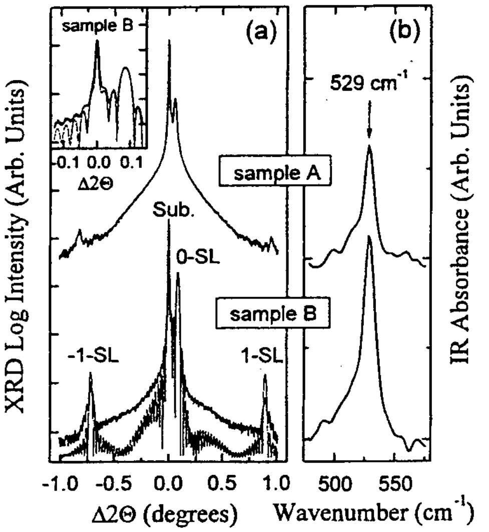


FIG. 2. Symmetric (004) x-ray diffraction rocking curves of $Ge_{1-y}C_y$/Ge superlattices with nominal C content of 1.2% (sample A) and 2.1% (sample B), respectively and best fitted simulated curve for sample B (a). The good crystalline quality is reflected in the observation of higher order SL peaks and the observable Pendelosung fringes around the zero-order peak [see also inset of (a)]. IR absorption of the two $Ge_{1-y}C_y$/Ge SL less a Ge reference spectra (b). The absorption line observed at 529 cm^{-1} is attributed to the local vibrational mode of substitutional C in Ge.

Infrared absorption measurement were performed at room temperature in the frequency range of 450–4000 cm^{-1}. An absorbance spectrum measured on a reference Ge substrate was subtracted from the sample absorbance spectra to remove any irrelevant features originating from the substrate. Figure 2(b) shows the infrared absorbance spectra of the two $Ge_{1-y}C_y$/Ge SL. We observe an absorption line at 529 cm^{-1} for both samples. Hoffmann *et al.*[12] studied substitutional C in Ge with infrared spectroscopy and ion channeling for Ge crystals implanted with C ions. They observed a local vibrational mode (LVM) due to substitutional C at 531 and 512 cm^{-1} for Ge crystals were ^{12}C and ^{13}C have been implanted, respectively. Therefore we attribute the absorption line at 529 cm^{-1} to the LVM of ^{12}C in Ge. The 2 cm^{-1} decrease in wavenumber is due to the different temperature during measurement.[13] The integrated absorbance of the mode at 529 cm^{-1} scales proportionally to the shift in the separation of the zero-order SL XRD peak with respect to the Ge substrate peak.

To investigate the thermal stability of the layers, the samples were annealed for 20 min at different temperatures from 300 to 850 °C. After the annealing, measurements of symmetric (004) and asymmetric (115) XRD rocking curves were carried out, to determine the perpendicular and in-plane lattice constant. The layers do not relax by nucleation of misfit dislocations since no change of the in-plane lattice constant is detected within this temperature range. Figure 3 shows the angle distance $\Delta 2\Theta$ between the main (004) peaks of the Ge substrate and the $Ge_{1-y}C_y$ layers. The data in Fig. 3 show that $\Delta 2\Theta$ does not change up to a temperature of 450 °C. At higher annealing temperatures $\Delta 2\Theta$ decreases, indicating the thermal activation of C atoms located in the Ge matrix. The temperature behavior is in agreement with IR absorption and ion channeling measurements by Hoffmann *et al.*[12] which are carried out in the same temperature range.

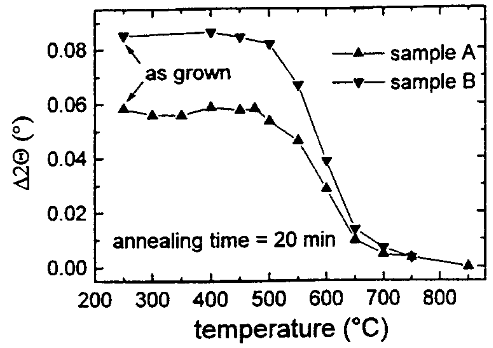


FIG. 3. Angle distance Δ2Θ between the main (004) peaks of the Ge substrate and the $Ge_{1-y}C_y$ layers as a function of annealing temperature. At temperatures above 450 °C Δ2Θ decreases indicating the thermal activation of substitutional C.

It indicates, that the films are likely to relax by C diffusion into precipitates and not by nucleation of misfit dislocations like in the case of SiGe/Si heterostructures. To investigate the loss of substitutional C in more detail the time dependence of the decrease of C in the SL structure is measured at different temperatures. The normalized angle distance $\Delta 2\Theta(t)/\Delta 2\Theta(t=0)$ in Fig. 4(a) shows that the loss of substitutional C is independent of the initial carbon content but increases with increasing annealing temperature. In the theory of phase transformations in alloys, this behavior can be described by the semiempirical rate equation $y(t) = \exp[(-k_r t)^n]$, where $y(t)$ is the fraction not transformed and k_r is the empirical rate constant.[14] The empirical parameter n characterizes the model process for the growth of the precipitates. The solid lines in Fig. 4(a) are best fitted simulated curves using n and k_r as fitting parameters. We find values of n=0.4–0.5 for the $Ge_{1-y}C_y$ structures, almost independent of the annealing temperature and initial carbon content. In the case of $Si_{1-y}C_y$ alloys, a value of $n=1$ is found for the diffusion controlled formation of the very stable SiC phase, which dominates the formation of precipitates.[15] In the Ge/C system the measured value of n=0.5 indicates a different nature of precipitate formation. These results are in agreement with Raman measurements, which indicate the formation of amorphous and graphitic carbon clusters in $Ge_{1-y}C_y$ alloys, while no stable GeC phase is known to exist.[7] An activation energy E_A for the loss of substitutional C can be obtained from the temperature dependence of the rate constant $k_r \sim \exp[-E_A/kT)]$. Figure 4(b) shows the Arrhenius plot $\log(k_r)$ against *1000/T* for both SL structures. The solid line is a linear fit to the data, which corresponds to an activation energy of E_A=2.6 eV. Kulik *et al.* report that the activation energy for the loss of substitutional carbon in $Si_{1-x-y}Ge_xC_y$ layers with a carbon content of about 0.8% has a maximum of E_A=4.9 eV for strain compensated $Si_{0.892}Ge_{0.1}C_{0.008}$ alloy on Si, and decreases for higher Ge contents.[16] This is explained by an increase of the mismatch between Si–Si and Si–C bonds with increasing Ge concentration. From their results, an activation energy below 3 eV can be assumed for Ge-rich $Si_{1-x-y}Ge_xC_y$ layers, which is in agreement with our measurement.

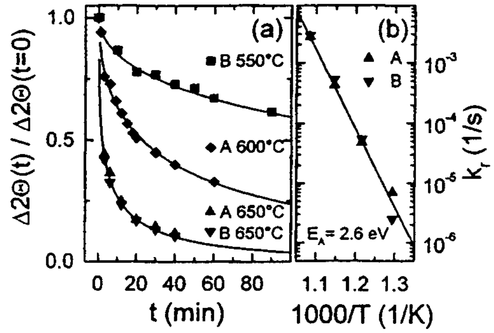


FIG. 4. (a) Normalized angle distance $\Delta 2\Theta(t)/\Delta 2\Theta(t=0)$ for different annealing temperatures. (b) Arrhenius plot of the rate constant k_r over *1000/T*. The straight line is a linear fit to the data.

In conclusion, we have shown that pseudomorphic $Ge_{1-y}C_y$ layers can be grown by MBE with carbon contents of up to 0.8%. The efficiency of substitutional C incorporation at a growth temperature of 200 °C is about 30%, which is considerably lower as compared to the nearly complete incorporation of C in Si. Narrow IR absorption at 529 cm^{-1} due to substitutional carbon in Ge is observed. Annealing experiments show the stability of $Ge_{1-y}C_y$ alloys up to a temperature of about 450 °C. At higher temperatures, an activation energy for the loss of substitutional carbon of E_A =2.6 eV is determined.

[1] K. Eberl, K. Brunner, and O. G. Schmidt, in *Germanium Silicon Science and Technology*, edited by R. Hull and J. C. Bean (Semiconductor and Semimetals, Germanium Silicon, 1998).

[2] K. Eberl, Festkörperprobleme **37**, 25 (1997).

[3] O. G. Schmidt and K. Eberl, Phys. Rev. Lett. **80**, 3396 (1998).

[4] B.-K. Yang, M. Krishnamurthy, and W. H. Weber, J. Appl. Phys. **84**, 2011 (1998).

[5] K. Brunner, K. Eberl, W. Winter, and E. Bugiel, Appl. Surf. Sci. **102**, 17 (1996).

[6] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, Appl. Phys. Lett. **67**, 1865 (1995).

[7] B.-K. Yang, M. Krishnamurthy, and W. H. Weber, J. Appl. Phys. **82**, 3287 (1997).

[8] H. J. Osten, E. Bugiel, and P. Zaumseil, J. Cryst. Growth **142**, 322 (1994).

[9] P. C. Kelires, Phys. Rev. B **55**, 8784 (1997).

[10] S. Zerlauth, H. Seyringer, C. Penn, and F. Schäffler, Appl. Phys. Lett. **71**, 3826 (1997).

[11] P. Zaumseil, Phys. Status Solidi A **141**, 155 (1994).

[12] L. Hoffmann, J. C. Bach, B. Bech Nielson, P. Leary, R. Jones, and S. Öberg, Phys. Rev. B **55**, 11167 (1997).

[13] R. C. Newman, Mater. Res. Soc. Symp. Proc. **59**, 403 (1986).

[14] R. J. Borg and G. J. Dienes, *The Physical Chemistry of Solids* (Academic, Boston, 1992), p. 521.

[15] G. G. Fischer, P. Zaumseil, E. Bugiel, and H. J. Osten, J. Appl. Phys. **77**, 1934 (1995).

[16] L. V. Kulik, D. A. Hits, M. W. Dashiell, and J. Kolodzey, Appl. Phys. Lett. **72**, 1972 (1998).

PHYSICAL REVIEW B VOLUME 59, NUMBER 24 15 JUNE 1999-II

Phonon spectra of substitutional carbon in $Si_{1-x}Ge_x$ alloys

L. V. Kulik
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716
and Institute of Solid State Physics, RAS, 142432 Chernogolovka, Moscow region, Russia

C. Guedj, M. W. Dashiell, and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

A. Hairie
Laboratoire d'Etudes et de Recherches sur les Materiaux (CNRS UA 6004), Institut des Sciences de La Matiere et du Rayonnement-6,
Boulevard du Marechal Juin, 14050 Caen Cedex, France

(Received 28 September 1998; revised manuscript received 19 January 1999)

The phonon spectra of substitutional carbon defects in $Si_{1-x}Ge_x$ alloys are studied over a broad range of Ge contents. With changing composition, the threefold-degenerate localized phonon mode of carbon defects in silicon was found to split into a number of phonon modes. The split-off modes appear due to the warping of defect symmetry while different nearest atomic environments of silicon and germanium atoms form around substitutional carbon atoms. The phonon spectra are simulated with a Keating model, taking into account the effects of carbon anharmonicity. The results of calculations are in good agreement with experiments. [S0163-1829(99)03724-8]

INTRODUCTION

Substitutional carbon in Si and SiGe alloys has attracted substantial interest due to the effects of carbon on the structural, optical, and electronic properties of these semiconductor materials.[1-5] Isolated substitutional and interstitial C defects in silicon have been thoroughly studied both experimentally and theoretically (for a review, see Ref. 6). Conversely, an extensive study of the carbon defect in SiGe alloys has commenced only recently, when nonequilibrium growth techniques such as molecular-beam epitaxy (MBE), chemical vapor deposition, and solid phase epitaxy (SPE) demonstrated carbon concentrations many orders of magnitude in excess of its equilibrium solubility.[7-11]

Because of the significant difference in covalent bond radii between Si and C atoms, the C defect in the Si lattice is surrounded by a considerable strain field, with substitutional defect bonds under tensile strain, and interstitial ones under compressive strain. The minimum formation energy of the interstitial C defect [in a split (100) configuration] exceeds the formation energy of the substitutional C defect by 3 eV,[12] which is why C atoms are favored to occupy substitutional sites.

A substitutional C defect in a SiGe alloy is expected to be surrounded by an even stronger strain field than in pure silicon, due to the increase of the C bond lengths with the Ge atoms. Interstitial C defects will be under smaller compressive strain due to the increase of the average lattice constant with Ge. Thus one could expect a decrease of the energy barrier between substitutional and interstitial C defects in $Si_{1-x}Ge_x$ alloys. The direct support for this was found in Ref. 13, where the thermal stability of substitutional carbon in $Si_{1-x-y}Ge_xC_y$ alloys was measured. The activation energy of substitutional C loss was shown to decrease by more than 1 eV in $Si_{0.7-y}Ge_{0.3}C_y$ alloys compared to $Si_{1-y}C_y$ alloy of similar y values.

Recently, the thermal stability of the substitutional C defect in germanium produced by the SPE technique was reported.[14] The highest obtained ratio of substitutional to total carbon concentration was only 0.3:1, while a similar technique applied to silicon resulted in nearly 100% carbon substitutionality.[15] The substitutional C defects in Ge were detected only in a narrow region of recrystallization temperatures which was far below the stability limit of substitutional carbon in silicon. Combined with previous observations that carbon solubility in germanium is 9–10 orders of magnitude less than in silicon,[16] this indicates that the formation energy of the substitutional C defect in germanium considerably exceeds that in silicon, and the energy barrier between substitutional and interstitial sites in Ge becomes relatively small.

Being of T_d symmetry, the isolated substitutional $C(^{12}C)$ defect in silicon is characterized by a localized threefold-degenerate phonon mode at 605 cm^{-1}, which is Raman and infrared active.[6] If the concentration of C defects becomes high, the interaction between the individual defects results in broadening of the localized phonon mode and weaker satellite phonon modes.[15,17] Intensities of the satellite modes reflect the distribution of the relative C arrangements through silicon, among which the arrangement of C atoms as the third neighbors was especially favorable.[17]

Because of enormous difficulties in diluting any amount of carbon available for optical studies in germanium, the phonon spectrum of substitutional C in germanium was not reported until recently.[14] The localized phonon mode of the isolated substitutional $C(^{12}C)$ defect in germanium has been observed at 531 cm^{-1}.

In the present paper we study the phonon spectra of the substitutional C defect in $Si_{1-x}Ge_x$ alloys, both experimentally using infrared-absorption spectroscopy, and theoretically using an anharmonic Keating model. The phonon frequencies of the substitutional C defect are measured vs Ge content. A number of phonon modes associated with the sub-

©1999 The American Physical Society

stitutional C defect are detected, and are attributed to localized phonon modes of C defects with different nearest-neighbor configurations of Si and Ge atoms.

EXPERIMENTAL PROCEDURE

The $Si_{1-x}Ge_xC$ alloys were grown by solid-source MBE on float zone Si(100) substrates. A combination Si-graphite source was used to form a molecular beam of pure Si and Si-C molecular species, resulting in a Si:C ratio of approximately 100:0.8 in the as-grown samples.[18] High-purity Ge was evaporated from a standard effusion source. The Si-graphite source temperature and growth time were held constant for all the samples maintaining a constant Si:C ratio and total number of carbon atoms in the layer. The Ge source temperature was varied to produce $Si_{1-x}Ge_xC$ alloys with increasing Ge content ($0 \leqslant x \leqslant 0.9$). The substrate temperature was about 200 °C.

The low substrate temperature was chosen to minimize the effects of near-surface diffusion of C atoms. Previous studies of the surface kinetics of C atoms during the MBE growth of $Si_{1-y}C_y$ alloys reveal a thermally activated diffusion of C atoms at the dimerized Si(100) surface from substitutional to near-surface interstitial sites, and conglomeration of C atoms in coherent and incoherent SiC precipitates.[18–20] A similar trend was found during the growth of $Si_{1-x}Ge_xC$ alloys, however, the activation energies and pre-exponential factors differed significantly. For increased Ge concentrations, lower growth temperatures were found to be necessary for a significant incorporation of C atoms in substitutional sites.

The as-grown samples were annealed at 450–750 °C for 30 min to crystallize the as-grown layer. Annealing temperatures were minimized to avoid the bulk diffusion of C atoms to sites other than random substitutional sites (such as the third-neighbor C arrangement mentioned above, and incoherent SiC precipitates). The annealing temperature depended on the stoichometry of the samples, as the activation energy for bulk diffusion of C atoms in $Si_{1-x}Ge_xC$ alloys is a decreasing function of Ge concentration.[13] The asymmetric (224) and symmetric (004) x-ray reflections were applied to quantify the Ge content in the alloy and the tetragonal distortion due to the interface strain between the alloy and substrate.[8,21] The alloys were found to be single crystalline and relaxed at $x > 0.2$. The alloy layers were pseudomorphic for $x \leqslant 0.1$. Alloy layer thickness and annealing temperatures for all values of x studied here are shown in Table I.

TABLE I. Layer thickness and annealing temperatures used to crystallize the $Si_{1-x-y}Ge_xC_y$ alloy layers. The layer thickness was varied so that the *total* number of carbon atoms in the layer remained constant for all values of x.

Ge composition (at. fraction)	Layer thickness (nm)	Recrystallization temperature (°C)
0	150	750
0.10	184	750
0.23	214	650
0.46	321	600
0.54	370	600
0.72	617	560
0.79	915	500
0.91	1900	475

The infrared-absorption measurements in the spectral region of the substitutional C vibration mode were performed at room temperature using a Nicolet 740 Fourier-transform infrared spectrometer with a Ga on KBr beam splitter operating in the transmission mode. A globar and a triglycerin sulfide detector were used as a light source and detector, respectively. The instrumental resolution was 1 cm^{-1}. The spectrometer had been purged with pure nitrogen for 1 h before the measurements to remove any environmental contribution to the IR spectra. Two problems were encountered during measurements: (i) a strong two-phonon absorption from the Si substrate in the same spectral region as the C-localized phonon mode; and (ii) interference fringes resulting from the difference in refractive index of the alloy layer and substrate, which complicates the spectra. While the standard differential method (subtraction of a reference substrate absorption spectrum from the sample spectrum) could successfully overcome the first of the above problem, it could not avoid the second for high Ge contents.

A relative differential method was applied in this study for high Ge contents, while the standard differential method was used for small Ge fractions of $x < 0.20$. The relative differential method is based upon the limited thermal stability of $Si_{1-x}Ge_xC$ alloys. Upon high-temperature treatment the alloy relaxes to its equilibrium state by the precipitation of β-SiC species of small grain size.[15,22] The β-SiC grains have a characteristic vibration above 800 cm^{-1} and do not contribute to the spectral region of the C-localized phonon mode. The IR spectra of the precipitated alloy sample (i.e., all substitutional C has transferred to incoherent β-SiC precipitates by a high-temperature anneal step) was used as a reference to be subtracted from the spectrum of the recrystallized sample spectra. Such a procedure was effective for removing all spectral features irrelevant to the substitutional C phonon modes, including interference fringes.

The studied samples were cut into 1-cm^2 pieces, measured and then annealed in a quartz furnace purged by pure nitrogen at atmospheric pressure in the temperature region from 700 to 1000 °C depending on the Ge content of the sample. The annealed samples were remeasured. Care was taken to guarantee precisely the same position of the sample in the fourier transform IR sample holder before and after the annealing. The operation was repeated until the associated C phonon modes disappeared, which was verified by a division of successively annealed sample spectra. The C phonon modes were considered to be negligible within the experimental error if the ratio of consecutive scans was unity for all wavelengths in the studied spectral range. The annealing time and temperature were chosen to minimize the Ge diffusion through the interface between the sample and substrate.

EXPERIMENT, THEORY, AND DISCUSSION

Figure 1 (bottom) shows a differential IR spectrum of a $Si_{1-y}C_y$ alloy in the spectral region of the localized phonon mode for C. The spectrum consists of a single band (T) associated with the localized phonon mode of the substitutional C defect, with no other spectral features observed. Upon a

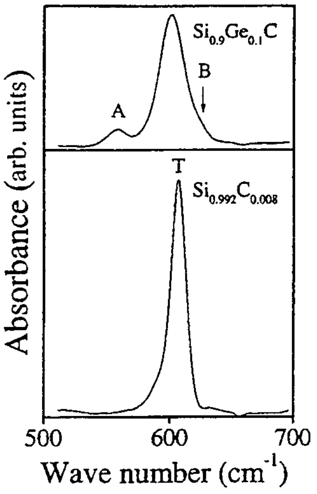


FIG. 1. Differential IR spectra of $Si_{0.992}C_{0.008}$ (bottom) and $S_{0.9}Ge_{0.1}C$ (top) recrystallized alloy samples. The labels *A, B*, and *T* are explained in the text.

small increase of Ge content [Fig. 1 (top)], the main absorption band shifts to lower frequencies and becomes significantly broader. In addition, a satellite band, *A* at 559 cm^{-1} and a shoulder *B* at 625 cm^{-1} become apparent. This trend is consistent with that observed by Hoffman[11] for relaxed $Si_{0.85}Ge_{0.15}$ alloys grown by MBE and implanted with $^{13}C^{+}$ ions for a concentration of $1 \times 10^{20} cm^{-3}$. For low Ge concentrations ($x<0.20$) they observed both low- and high-frequency absorption lines on either side of the main absorption line of the Si_4 vibration.

The origin of the *A* band was established by investigating the thermal stability of $Si_{0.9}Ge_{0.1}C$ alloys. Figure 2 shows the room-temperature differential IR spectra of the alloy subsequently annealed isothermally at 908 °C for different times. The spectra exhibit absorption bands at 602 and 815 cm^{-1}, and the satellite *A* band at 559 cm^{-1}. The *T* band at 602 cm^{-1} is the local phonon mode of the substitutional C defect, and the broad asymmetric band at 815 cm^{-1} is associated with the vibration of incoherent SiC precipitates.[15,22] During the annealing sequence, the absorbance of the *T* band decreases and that of SiC precipitates increases, showing that the C atoms transfer from substitutional sites to SiC precipitates.

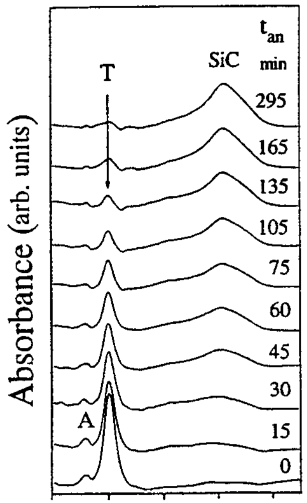


FIG. 2. IR absorbance spectra of the annealing sequence of a $Si_{0.9}Ge_{0.1}C$ alloy sample at 908 °C. Annealing times (t_{an}) are shown in the right side of the figure.

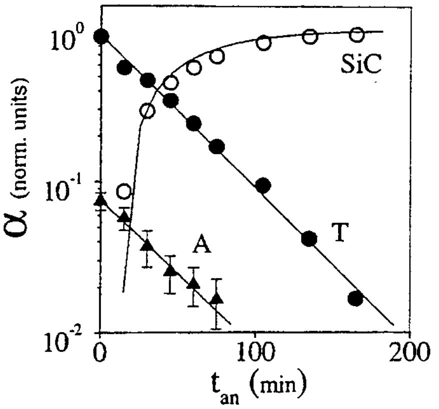


FIG. 3. The integrated IR absorbance (α) vs annealing time of the *T* (solid circles), *A* (triangles), and SiC precipitate (open circles) bands. Solid lines are the best fits of the experimental points.

As C atoms leave substitutional sites, the alloy, being initially strain compensated, gradually becomes compressively strained. We did not observe a frequency shift of the *T* band while the layer strain increased (see Fig. 2). Considering that the C content in $Si_{1-x}Ge_xC$, with $x>0.1$, decreases with increasing x (i.e., the Si:C ratio is constant), we conclude that the partial strain compensation of SiGe layers due to the addition of small fractions of carbon does not observably affect the phonon frequencies of the substitutional C defect.

The integrated absorbance of both *T* and *A* bands versus annealing time is shown in Fig. 3 on a semilogarithmic scale. Also shown is the increase of the absorbance of the broad band centered at 815 cm^{-1}. The decrease of *T* and *A* bands is fit with an exponential decay function:

$$\alpha_{T,A} = \alpha_0 \exp\left\{ -\frac{t}{\tau} \right\}, \qquad (1)$$

and the increase of the SiC precipitates band with

$$\alpha_{\text{precip}} = \alpha_1 \left(1 - \exp\left\{ -\frac{t}{\tau} \right\} \right). \qquad (2)$$

The time constant τ for all three bands is about 45 min within experimental error. As the formation of SiC precipitates is increased by the transfer of C atoms from substitutional sites, we interpreted the equality of the time constants for the growth of the SiC band and the decay of the *A* band as indications of a connection between the *A* band and a C defect. At the same time, having the same temperature stability as the *T* band, *A* is probably associated with a substitution C defect. In favor of this is the fact that interstitial C defects could not possibly account for the *A* band because

TABLE II. Table of Keating parameters used in simulations given in atomic units.

Bond	a_0	α	β
C-C	6.7424	0.143	0.101
Si-Si	10.266 32	0.0637	0.017
Ge-Ge	10.695	0.0555	0.015
Si-C	8.24	0.0555	0.0802
Si-Ge	10.4572	0.053	0.017
C-Ge	8.54	0.0742	0.031

they have a lower temperature stability[2] than substitutional C defects, and rather higher vibration frequencies.[23]

As the *A* band was not present in the spectrum of the sample without germanium, we attributed this to a phonon mode of a substitutional C defect with one Ge atom involved as the nearest-neighbor (Si_3Ge_1 configuration). This defect holds C_{3V} symmetry, and hence two absorption bands active in IR spectra are to be observed. The possibility of any spectral features relevant to configurations with more than one Ge nearest neighbor should be disregarded at small Ge contents due to the low probability for such configurations to form.

The validity of this assignment was tested by theoretical simulation. We applied a theoretical approach based upon a valence-force-field model derived from Keating, taking into account the effects of carbon anharmonicity. The interactions between atoms have been modeled with an interatomic potential similar to the one of Ref. 24, with the exception of adjusting our force coefficients to yield the correct lattice parameters and phonon modes of silicon, germanium, diamond (in the $Fd3m$ structure), SiGe and 3C-SiC at 300 K. This choice is justified by the very high precision and reliability of these experimental data, which can be obtained by x-ray diffraction, Raman spectroscopy, and absorption spectroscopy at room temperature. Table II displays the equilibrium lattice constants and Keating parameters determined from a least-square fit to the lattice constants and zone-center optical phonons of the corresponding diamond and zinc-blende structures. This set of parameters enables the computation of the LO (Γ) mode from 3C-SiC, which cannot be obtained from the parameters given in Ref. 24 to the best of our knowledge. Our model is therefore ideally suited to a precise computation of the lattice parameters and phonon spectra of tetrahedrally coordinated $Si_{1-x}Ge_xC$ alloys. The molecular-dynamics relaxation is calculated from a 512-atom supercell at time increments of 3 fs until the total potential energy reaches a stable minimum (approximately 2000 fs). We have computed isotropic relaxations, where no external pressure is applied to the computational box, to simulate a fully relaxed layer on silicon. In a further stage, the local phonon density around carbon is computed, using the recursion method detailed in Ref. 25. We have chosen to compute the local phonon spectra around carbon atoms situated near the center of the supercell to avoid surface effects. For each Ge concentration, several Si_aGe_b configurations of the carbon nearest neighbors were tested so that a particular configuration was held fixed, and the remaining atoms in the computation box were distributed randomly. The spectra were averaged over 16 random distributions.

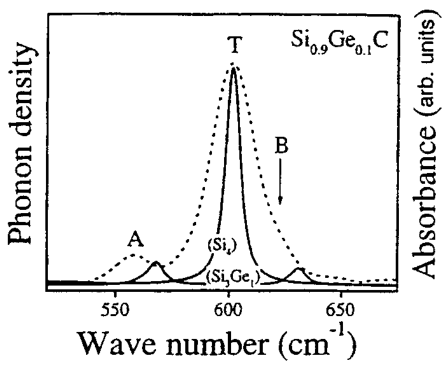


FIG. 4. Simulated localized phonon density around C (solid line), together with the measured IR absorbance spectrum (dots) for $x=0.1$. The relative populations if the two configurations Si_4 and Si_3Ge_t were determined by assuming that the probability to find a Ge or Si atom next to C is equal.

The simulated localized phonon density around C, together with the measured IR spectrum for $x \sim 0.1$ is shown in Fig. 4.[26] The simulation yielded the mode at *T* for the local vibrational mode (LVM) of C in Si_4. The simulated *A* and *B* modes are due to the low-frequency nondegenerate and high-frequency doubly degenerate phonon mode of the Si_3Ge_1 configuration, respectively. The phonon frequency of the substitutional C defect with the four Si nearest neighbors (Si_4) is close to the experimental position of the *T*-band maximum. The lower-frequency nondegenerate phonon mode of the experimental Si_3Ge_1 defect slightly exceeds (by 8 cm^{-1}) the frequency of the *A*-band maximum. We did not experimentally observe a separately resolved absorption band corresponding to the simulated high-frequency twofold-degenerate phonon mode of the Si_3Ge_1 defect. It may be that the frequency of this mode partly overlaps the frequency of the Si_4 defect phonon mode, and thus accounts for the observed high frequency shoulder of the *T* band, *B*, at 625 cm^{-1} (Fig. 4).

Upon a further increase of the Ge content, the effects of interference fringes on the IR spectra of $Si_{1-x}Ge_xC$ alloys become considerable, and we could no longer use the standard differential technique. As an example, the differential spectra of a $Si_{1-x}Ge_xC$ ($x \sim 0.79$) alloy grown on a Si substrate before (dots) and after long-time annealing (dashed line) are shown in Fig. 5. The differential spectra are not straightforward for interpretation due to interference fringes; however, a relative differential spectrum (solid line), which is a difference between the recrystallized and annealed spectra, is well defined.

Figure 6 shows a set of relative differential spectra of $Si_{1-x}Ge_xC$ alloys at $0 \leqslant x \leqslant 0.54$. The spectra are characterized by two absorption bands *T* and *A*, shifting linearly to low frequencies upon increase of the Ge content, with the slopes of -0.53 ± 0.3 and $-0.56 \pm 0.4\,cm^{-1}$ per 1% of germanium, respectively. The shift of the *T* band is in close agreement with the simulations, which give linear slope of -0.57 cm^{-1} per 1% of germanium.

Because of the C-C interaction, the full width at half maximum (FWHM) of the *T* band at $x=0$ is as much as 12 cm^{-1}, which is twice that of the individual substitutional C defect in silicon (6 cm^{-1}). At $x=0.1$, the FWHM of the *T* band becomes even broader ($\sim$20 cm^{-1}). This broadening is

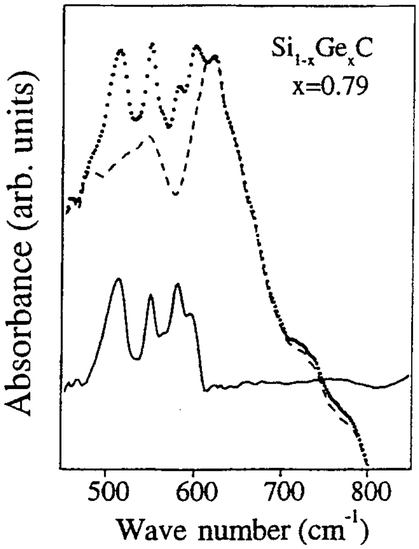


FIG. 5. Differential spectra of a $Si_{1-x}Ge_xC$ alloy for $x\sim0.79$ before (dots) and after long-time annealing (dashed line). The relative differential spectrum is shown by the solid line.

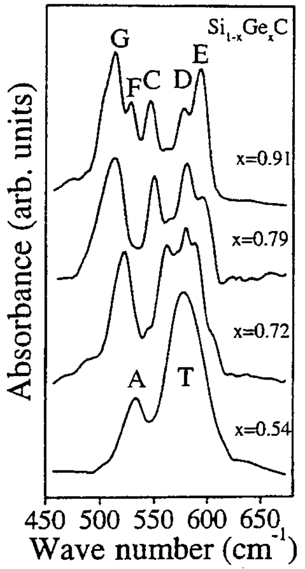


FIG. 7. Relative differential absorbance spectra of $Si_{1-x}Ge_xC$ alloys for $0.54\leq x\leq0.91$.

accounted by the effects of different *second*-neighbor, *third*-neighbor, etc. configurations, but not the *first*-neighbor configuration on the phonon frequency of the substitutional C defect in the alloy. It is not therefore surprising that the FWHM of the *A* band is of the same value, as the effect of remote neighbors is similar for substitutional C defects in both Si_4 and Si_3Ge_1 configurations.

The expected behavior of the *T*- and *A*-band FWHM's with increasing Ge content is determined by two opposing tendencies:(1) broadening due to the different second-neighbor, etc. configurations to achieve a maximum around $x\sim0.5$, and (2) narrowing due to reduction of the C-C interaction (our growth method preserves the SiC ratio, hence effectively increasing the average distance between C atoms in the alloys with increasing Ge content). The FWHM of the *A* band achieves a maximum at $x\sim0.3$, and then starts decreasing, showing that the second tendency indeed becomes dominating, see the inset of Fig. 6.

The *T*-band FWHM, however, behaves differently than expected. Having been equal to that of *A* band up to $x=0.23$, the FWHM starts increasing and achieves as much as 36 cm^{-1} at $x\sim0.54$ (the inset to Fig. 6). The only possible explanation for this huge broadening is that the *T* band no longer represents the phonon mode of substitutional C defect in one particular configuration, but appears to be a composite band. This becomes obvious from Fig. 7, where the relative

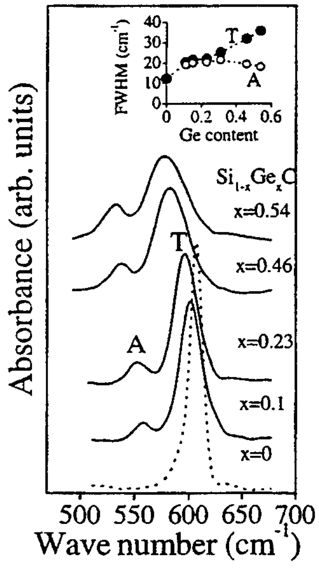


FIG. 6. Relative differential IR absorbance spectra of $Si_{1-x}Ge_xC$ alloys for $0\leq x\leq0.54$. In the inset, the full width at half maximum of the *T* and *A* bands are shown.

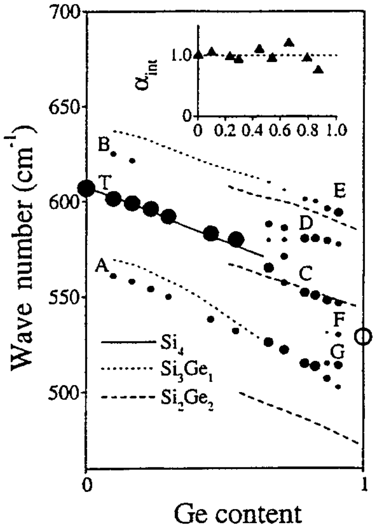


FIG. 8. Frequencies of the observed IR-absorption bands (solid circles) labeled in Fig. 7 vs Ge content. The dot areas reflect the integrated absorbance of each band determined by separately fitting each line to a Gaussian, followed by integration of the Gaussian fit. The lines are simulated frequencies of the localized phonon modes of substitutional C defects in particular nearest-neighbor configurations. The open circle at $x=1$ is the localized phonon mode frequency of the substitutional C defect in germanium. The inset shows the integrated absorbance through all IR-absorption bands associated with the substitutional C defect.

differential IR spectra of $Si_{1-x}Ge_xC$ alloys at higher Ge contents are shown.

The broad T absorption band splits into a number of narrow ones at $x>0.6$. At first, two bands, C and D, develop in the IR spectra (Fig. 7). Their relative absorbance stays nearly constant with increasing Ge content, which points out that both bands are associated with two phonon modes of the same defect. Having considered that the next statistically favorable substitutional C defect after the Si_3Ge_1 configuration has to be the Si_2Ge_2 one, we concluded that the C and D bands were associated with the two phonon modes of the substitutional C defect in the Si_2Ge_2 configuration. It holds a C_{2V} symmetry; hence three IR-active phonon modes are to be observed. The third expected mode has not yet been detected. It is quite possible that an absorption band corresponding to the third phonon mode is not resolved from the A band. In fact, the A band expands nearly twice in the narrow region of Ge contents, $0.54<x<0.7$, which may be interpreted as the contribution of two absorption bands associated with the substitutional C defect in the Si_3Ge_1 and Si_2Ge_2 configurations, overlapping in the same spectral region. Simulations made for the C defect in the Si_2Ge_2 configuration demonstrate that the two calculated phonon modes of the Si_2Ge_2 defect are best matched to the C and D bands, but the third phonon mode frequency situates noticeably below the frequency of A band.

At the highest Ge contents, an absorption band F develops in the IR spectra, with a frequency, if approximated to $x=1.0$, that tends to the frequency of the localized C phonon mode in pure germanium (Figs. 7 and 8). It was therefore attributed to the localized phonon mode of the substitutional C defect in the Ge_4 configuration. The rest of the observed absorption bands (E and G) dominating the IR spectra at the highest Ge contents, $x>0.8$, were tentatively assigned to the two IR-active phonon modes of the Si_1Ge_3 defect (Figs. 7 and 8).

Along with the frequencies of the observed absorption bands, we studied the total integrated absorption coefficient of all the absorption bands associated with substitutional C defects as a function of Ge content, shown in the inset to Fig. 8. Within experimental error, it is almost preserved through all Ge contents. This remarkable fact directly confirms the conclusion of Ref. 14, that the effective charge of the substitutional C atom in germanium has a similar value as in silicon. Moreover, this conclusion is now extended to $Si_{1-x}Ge_x$ alloys correspondingly. It could only be so if the effective charge of substitutional C in $Si_{1-x}Ge_x$ alloys does not significantly depend on the C-atom neighbors.

CONCLUSION

The phonon spectra of substitutional C defects in $Si_{1-x}Ge_x$ alloys are studied both experimentally and theoretically. The effect of symmetry warping due to different *first*-neighbor, *second*-neighbor, etc. configurations of Si and Ge atoms around substitutional carbon on the phonon spectra of the defect is discussed. It is found that the effective charge of the substitutional C atom in $Si_{1-x}Ge_x$ is not affected significantly by the C-atom neighbors. The theoretical model developed in Ref. 19 for the simulation of phonon spectra of $Si_{1-x}Ge_xC$ alloys is tested for broad regions of Ge contents. Close agreement between experimental and theoretical results is obtained.

ACKNOWLEDGMENTS

This work was supported by the Army Research Office under Grant No. DAAH04-95-1-0625 and Grant No. AA-SERT DAAG55-97-1-0249, by the Defense Advanced Research Projects Agency under Contract No. F49620-96-C-0006, and by the Office of Naval Research under Grant No. N00014-93-1-0393.

---

[1] G. Davies and R. C. Newman, in *Handbook on Semiconductors*, edited by T. S. Moss (Elsevier Science, Amsterdam, 1994), Vol. 3b, p. 1557, and references therein.

[2] G. D. Watkins and K. L. Brower, Phys. Rev. Lett. **36**, 1329 (1976).

[3] L. W. Song, X. D. Zhan, B. W. Benson, and G. D. Watkins, Phys. Rev. Lett. **60**, 460 (1988).

[4] O. F. Sankey, A. A. Demkov, W. T. Petuskey, and P. F. McMillan, Modell. Simul. Mater. Sci. Eng. **1**, 741 (1993).

[5] P. C. Kelires, Phys. Rev. Lett. **75**, 1114 (1995).

[6] R. C. Newman, Adv. Phys. **18**, 544 (1969).

[7] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, Appl. Phys. Lett. **60**, 3035 (1992).

[8] B. Dietrich, H. J. Osten, H. Rücker, M. Methfessel, and P. Zaumseil, Phys. Rev. B **49**, 17 185 (1994).

[9] A. St. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewolt, Appl. Phys. Lett. **67**, 3915 (1995).

[10] J. W. Strane, H. J. Stein, S. R. Lee, B. L. Doyle, S. T. Picraux, and J. M. Mayer, Appl. Phys. Lett. **63**, 2786 (1993).

[11] L. Hoffman, J. C. Bach, J. L. Hansen, A. N. Larsen, B. B. Nielson, P. Leary, R. Jones, and S. Oberg, in *Defects in Semiconductors ICDS-19*, edited by G. Davies and M. H. Nazare, [Mater. Sci. Forum **258–263**, 97 (1997)].

[12] J. Tersoff, Phys. Rev. Lett. **74**, 5080 (1995).

[13] L. V. Kulik, D. A. Hits, M. W. Dashiell, and J. Kolodzey, Appl. Phys. Lett. **72**, 1972 (1998).

[14] L. Hoffman, J. C. Bach, B. Bech Nielsen, P. Leary, R. Jones, and S. Öberg, Phys. Rev. B **55**, 11 167 (1997).

[15] J. W. Strain, H. J. Stein, S. R. Lee, S. T. Picraux, J. K. Watanabe, and J. W. Mayer, J. Appl. Phys. **76**, 3656 (1994).

[16] R. I. Scace and G. A. Slack, J. Chem. Phys. **30**, 1551 (1959).

[17] H. Rücker, M. Methfessel, B. Dietrich, K. Pressel, and H. J. Osten, Phys. Rev. B **53**, 1302 (1996).

[18] M. W. Dashiell, L. V. Kulik, D. A. Hits, J. Kolodzey, and G. Watson, Appl. Phys. Lett. **72**, 833 (1998).

[19] C. Guedj, M. W. Dashiell, L. V. Kulik, J. Kolodzey, and A. Hairie, J. Appl. Phys. **84**, 4631 (1998).

[20] H. J. Osten, E. Bugiel, and P. Zaumseil, Appl. Phys. Lett. **64**, 3440 (1994).

[21] M. Fatemi and R. E. Stahlbush, Appl. Phys. Lett. **58**, 825 (1991).

[22] J. Mi, P. Warren, P. Letourneau, M. Judelewicz, M. Gailhanou,

M. Dutoit, C. Dubois, and J. C. Dupuy, Appl. Phys. Lett. **67**, 259 (1995).

[23] A. R. Bean and R. C. Newman, Solid State Commun. **8**, 175 (1970).

[24] H. Rücker and M. Methfessel, Phys. Rev. B **52**, 11 059 (1995).

[25] A. Hairie, F. Hairie, G. Nouet, E. Paumier, and A. P. Sutton, in *Polycrystalline Semiconductors III–Physics and Technology*, edited by H. P. Strunk, J. H. Werner, B. Fortin, and O. Bonnaud (Trans Tech, Zurich, 1993). Vols. 37 and 38, p. 91.

[26] The calculated phonon densities give only a rough measure of the intensities of the absorption lines associated with local vibrational modes.

Molecular beam epitaxy growth of $Ge_{1-y}C_y$ alloys on Si (100) with high carbon contents

K. J. Roe,[a] M. W. Dashiell, and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716
P. Boucaud and J.-M. Lourtioz
Institut d'Electronique Fondamentale, Université Paris Sud, 91405 Orsay, France

(Received 30 October 1998; accepted 23 March 1999)

Group IV alloys are attracting strong interest for Si-based optoelectronics. The effects of C on the electrical and optical properties, however, are still not well understood, especially for high Ge content. In this report, we describe optical, structural, and compositional measurements of a series of thick, relaxed *p*-type $Ge_{1-y}C_y$ layers on *n*-type Si (100) substrates. The alloy layers were 0.5 μm thick and were grown by solid source molecular beam epitaxy at a substrate temperature of 300 °C and *p*-type doped with different B concentrations. X-ray diffraction indicated that the layers were single crystalline and nearly fully relaxed. The optical absorption was measured using a waveguide structure using Fourier transform infrared spectroscopy. The absorption data versus photon energy data fit indicated an indirect band gap, and one sample had a band gap of 774 meV compared to 660 meV for pure Ge. For single-crystalline, relaxed layers, the effect of C was to increase the band gap energy. These measurements show that alloying Ge with C provides a way to vary the optical absorption, which may be useful for device applications. © *1999 American Vacuum Society.* [S0734-211X(99)07103-6]

Group IV alloys are attracting attention for use in Si-based optoelectronics. Optical and electrical devices use $Si_{1-x}Ge_x$ alloys to improve device performance compared to pure Si, and C may give new possibilities for devices. For example, it has been shown that the addition of C stabilizes the alloys by reducing strain and dopant out-diffusion.[1] Carbon has a low equilibrium solid solubility in Ge, so therefore a nonequilibrium growth technique such as molecular beam epitaxy (MBE) is necessary to achieve significant C concentrations. Metastable $Ge_{1-y}C_y$ alloys with carbon concentrations of $y \approx 0.01$ have been grown by MBE at low growth temperatures, near 400 °C.[2] Here we report on the growth, optical, and structural properties of crystalline $Ge_{1-y}C_y$ alloys with a range of C content.

The $Ge_{1-y}C_y$ alloys were grown by solid-source MBE in an EPI Model 620 system. The system features six effusion cells and a substrate introduction chamber. The typical base pressure of the system before growth was on the order of 10^{-11} Torr. During growth, the system pressure was typically 5×10^{-9} Torr, maintained by a liquid helium cooled cryopump.

The Ge beam was formed by thermally evaporating triple zone-refined intrinsic Ge in a pyrolytic boron nitride (PBN) crucible. To minimize boron contamination from the crucible, the cell temperature was kept below 1380 °C. At 1325 °C the Ge growth rate was approximately 0.2 Å/s.

The C beam was produced by an EPI single-filament carbon source. The beam is created by sublimating a pyrolytic graphite filament resistively heated by a direct current power supply. Typical current values were 48–50 A, resulting in an estimated filament temperature of 2300 °C.

Substrates were *p*-type (100)-oriented 3 in. diameter single crystal Si wafers with a resistivity of 1–10 Ω cm. Wafers were prepared by degreasing, etching in $H_2O:H_2O_2:HCl(5:3:3)$, and dipping in $HF:H_2O(1:10)$ to terminate the surface with H.[3] Wafers were then immediately loaded into the MBE for growth. The substrate was then heated above 200 °C for 1 h to desorb any surface contaminants. The growth conditions for the samples are shown in Table I. Alloys were grown at a substrate temperature of 300 °C. Substrate temperature values of a nearby thermocouple were calibrated by the Si eutectic transition with Al and Au. Alloys were grown at low temperature in an attempt to maximize the substitutional C incorporation. The GeC layer was grown directly onto the Si substrate without any buffer layers. The aim was to grow thick relaxed films in order to study the bulk properties of unstrained layers, which were thick enough so that the interface dislocations would not adversely affect the measured parameters. Reflection high-energy electron diffraction (RHEED) analysis both during and after growth showed 2×1 surface reconstruction. Layer thicknesses were estimated from growth conditions and confirmed by Rutherford backscattering (RBS) to be 0.5 μm.

Optical absorption measurements were performed by Fourier transform infrared (FTIR) spectroscopy measurements. A waveguide reflection method was used to obtain approximately eight bounces through the alloy film.[4] Figure 1 illustrates the waveguide reflection geometry. The samples were prepared to a size of 5 mm×8 mm and mechanically polished at a 45° angle at opposite ends. FTIR data from a pure Si waveguide reference was subtracted from the waveguide data to yield the absorption of the layers alone. This layer

---

[a] Electronic mail: roe@ee.udel.edu

TABLE I. MBE growth conditions for the GeC layers; the substrate temperature was 300 °C and the Ge cell temperature was 1325 °C for all samples.

Sample number	C cell current (A)	B cell temp. (°C)
SGC-261	49	1550
SGC-266	48.5	1550
SGC-267	49.5	1550
SGC-270	50	1550
SGC-102	0	1450
SGC-103	0	1650

absorption was not calibrated for possible optical scattering or leakage during the multiple reflections.

The absorption data were fit to the McFarlane–Roberts expression for band gap, E_g and phonon energy, E_{ph} above the band gap

$$\alpha = A\frac{(h\nu - E_g + E_{ph})^2}{e^{E_{ph}/kT}-1}, \quad E_g - E_{ph} < h\nu < E_g + E_{ph},$$

$$\alpha = A\left[\frac{(h\nu - E_g - E_{ph})^2}{1-e^{-E_{ph}/kT}} + \frac{(h\nu - E_g + E_{ph})^2}{e^{E_{ph}/kT}-1}\right], \quad h\nu > E_g + E_{ph},$$

where $h\nu$ is the photon energy, E_g is the band gap, and E_{ph} is the phonon energy. Figure 2 shows the measured data and the MacFarlane–Roberts expression fit. Table II lists the measured optical band gaps for each of the samples. For our thick samples, the effect of alloying was to increase the band gap. For sample SGC-261, we observe a band gap of 772 meV, which is more than 100 meV greater than bulk Ge (664 eV). This large shift in band gap is useful for GeC/Ge heterostructures. We attribute the band gap difference to C incorporation, noting that diamond has a band gap of 5.45 eV. These layers are near fully relaxed, and thus the strain in these samples is small. Thus bulk strain reduction cannot account for all of the change in observed band gap. If we assume Vegard's Law, however, and attribute the entire shift in band gap to carbon incorporation, we infer C concentrations of $y \approx 0.02$. Since the relation between band gap and C fraction is not yet known, this linear method is not reliable for determining C concentration, especially for large C concentrations. It is known that the band gap of 3C-SiC($E_g = 2.2\,\mathrm{eV}$) is much less than predicted by Vegard's Law ($E_g = 3.28\,\mathrm{eV}$) by interpolating between diamond and Si.

X-ray diffraction (XRD) measurements were performed to determine the structure, relaxation, and concentration of the films. The XRD measurements were performed on a Philips X'Pert MRD materials research diffractometer using a Cu$K\alpha$ radiation source. The Cu$K\alpha$, radiation is selected by a four-bounce Ge (220) Bartels monochromator, and a triple-axis output monochromator for high resolution. The system is also equipped with a fully rotational rocking curve stage, which allows measurement of off-axis reflections. This allows the measurement of the d spacings and lattice parameters in the direction ($a_\perp$) of growth and in the plane of the substrate ($a_\parallel$). From these reflections, we can infer tilt, relaxation, and tetragonal distortion. Using the Poisson ratio, the bulk lattice parameter (a_0) can be inferred. In each case, the layer reflections are referenced to the Si substrate peak present in each scan. The substrate peak is corrected to the nominal value for Si($a = 5.430\,88$ Å). Any correction in 2Θ to the measured Si peak is then applied to the 2Θ value of the layer peak. For the analysis in this article, the layers are assumed to exhibit only simple tetragonal strain, although there is the possibility of monoclinic or triclinic distortion.[5] The specific nature of the strain of these layers was not meant to be addressed in this work.

Wide-range scans along the surface normal showed only (004) reflections from the layers, indicating the alloy layers have a structure highly oriented to the (100) substrate. The full width at half maximum (FWHM) values for the layers were on the order of tenths of a degree in 2Θ, indicating good quality layers. For this work, the (004), (113), and (224) + and − reflections were measured for each sample. The + and − designations refer to the opposite sample orientations in the diffraction plane, which, when averaged, give us the corrected scattering angle, 2Θ.[6] Figure 3 shows the (004)+ scan of one of our samples. We calculate $a_\perp$ from the average of the two (004) +/− reflections. We calculate $a_\parallel$ from the (113) and (224) +/− reflections by first

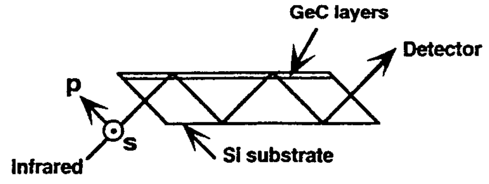


FIG. 1. Waveguide FTIR geometry for multiple bounces through the film. Electric polarization orientations (s and p) are indicated.

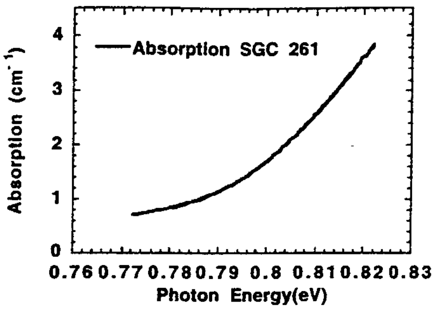


FIG. 2. FTIR absorption data for sample SGC-261 (thin line). The data are fit to the MacFarlane–Roberts expression above the band gap. Note how well the model (dark line) fits the data for an indirect band gap.

TABLE II. Measured optical band gaps for the GeC layers.

Sample number	Optical band gap (meV)
SGC-261	772.18
SGC-266	766.66
SGC-267	763.13
SGC-270	766.31
bulk Ge	664.00

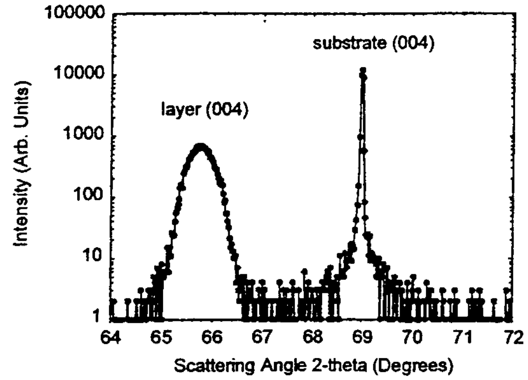


FIG. 3. XRD diffraction intensity vs diffraction angle of (004) layer/substrate reflection of sample SGC-261. The GeC layer peak appears around 65.8°, and the Si substrate peak occurs around 69°. Note the small width of the Si substrate peak, indicating a sharp system response.

using geometry to calculate the parallel lattice parameter for each reflection, and then averaging the four values.

The bulk lattice parameter (a_0) was also then calculated from $a_\perp$ and $a_\parallel$ as follows[7]:

$$a_0 = a_\perp + \frac{[2v(a_\parallel - a_\perp)]}{1+v},$$

where v is the linear interpolated Poisson ratio for GeC from the values for Ge and C.[8] The layer relaxation parameter, R, is then

$$R = \frac{a_\parallel - a_{Si}}{a_0 - a_{Si}},$$

where a_{Si} is the substrate lattice parameter.[7] We assume Vegard's Law holds for the lattice parameter and back out an estimate for the C concentration (y). As references, we grew samples with no C [pure Ge on (100) Si], and then attributed the change in bulk lattice parameter in the alloys to the C. We then used the range of values obtained from the reference samples to obtain an average C concentration and error. The instrument error in XRD measurements is very small, especially with a high-resolution instrument. In addition, a near perfect peak fit using a Lorentzian or Pearson VII function (with confidence factor >0.95) is obtained for the measured XRD data.[2] Table III shows values for a_0, relaxation, and the C concentration.

TABLE III. XRD data and calculations for the GeC layers.

Sample number	Bulk lattice parameter, a_0 (A)	Relaxation (R)	Calculated carbon concentration
SGC-261	5.6482	0.9022	0.64%±0.16
SGC-266	5.6491	0.9155	0.59%±0.16
SGC-267	5.6571	0.9778	0.39%±0.16
SGC-270	5.6533	0.9439	0.21%±0.16
SGC-102	5.6582	0.9537	0%[a]
SGC-103	5.6650	0.9168	0%[b]

[a] Reference 1.
[b] Reference 2.

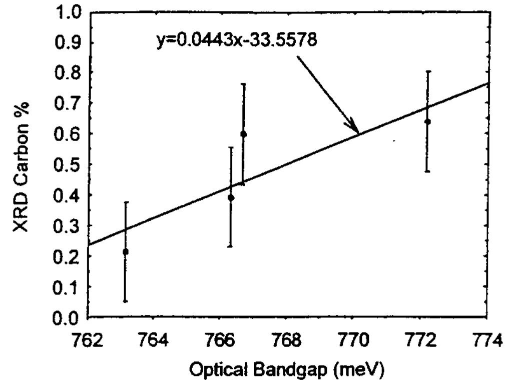


FIG. 4. Calculated carbon fraction vs FTIR measured optical band gap. The least-squares best-fit line is plotted. Note the clear trend of increasing band gap with carbon concentration.

Comparing the FTIR band gap and XRD calculated C concentration yields an interesting result. Figure 4 shows a plot of band gap versus calculated C concentration. We find that C *increases* the optical band gap of Ge on Si. The least-squares best-fit line has been plotted on the graph as a convenience. We expect due to X and L maxima mixing, that the true relationship has some bowing. The best-fit line prediction of $E_g = 3.02$ eV underestimates the C diamond band gap ($E_g = 5.45$ eV).[8] However, the trend of increased band gap with increasing C concentration is clear.

In conclusion, we have grown crystalline GeC alloys using low-temperature MBE. Measurements on the alloys layers show the incorporation of C increases the Ge band gap by more than 100 meV. XRD measurements confirm the crystalline structure of the samples and estimate C concentrations to be 0.2%–0.6%. These alloys offer a new material for use in Group IV heterostructures.

The authors gratefully acknowledge T. Tröger and T. Adam for useful discussion, and P. Thompson and K. Hobart for advice on MBE growth. This work was funded by the U.S. Army Research Office, Grant No. DAAH04-95-1-0625, and the Office of Naval Research, Grant No. N00014-93-1-0393.

[1] H. J. Osten, B. Heinemann, D. Knoll, G. Lippert, and H. Rücker, J. Vac. Sci. Technol. B **16**, 1750 (1998).
[2] J. Kolodzey, P. A. O'Neil, S. Zhang, B. A. Orner, K. Roe, K. M. Unruh, C. P. Swann, M. M. Waite, and S. Ismat Shah, Appl. Phys. Lett. **67**, 1865 (1995).
[3] B. A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J. W. Mayer, and J. Spear, Appl. Phys. Lett. **69**, 2557 (1996).
[4] S. Sauvage, P. Boucaud, F. H. Julien, J.-M. Gerard, and J.-Y. Marzin, J. Appl. Phys. **82**, 3396 (1997).
[5] W. J. Bartels, J. Vac. Sci. Technol. B **1**, 338 (1983).
[6] M. Fatemi and R. E. Stahlbush, Appl. Phys. Lett. **58**, 825 (1991).
[7] M. A. G. Halliwell, Appl. Phys. A: Solids Surf. **58**, 135 (1994).
[8] *Data in Science and Technology: Semiconductors, Group IV Elements and III–V Compounds*, edited by O. Madelung (Springer, Berlin, 1991).

Appendix 45

PHYSICAL REVIEW B VOLUME 60, NUMBER 22 1 DECEMBER 1999-II

Structure and lattice dynamics of $Ge_{1-y}C_y$ alloys using anharmonic Keating modeling

C. Guedj* and J. Kolodzey
Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716

A. Hairie
Laboratoire d'Etudes et de Recherches sur les Matériaux, Institut des Sciences de la Matière et du Rayonnement, UPRESA-CNRS 6004, 6 Boulevard du Maréchal Juin, 14050 Caen Cedex, France

(Received 21 December 1998; revised manuscript received 20 August 1999)

The incorporation of substitutional C into Ge is studied theoretically with an anharmonic Keating model fitted to recent experimental data and specifically adapted to the computation of the structural properties and the lattice dynamics of $Ge_{1-y}C_y$ alloys. In the range of physically realistic concentrations ($y<3\%$), the change of lattice parameter due to substitutional carbon is found to agree with Vegard's law of linear interpolation between germanium and diamond with a maximum relative deviation less than 0.03%. This result is obtained in the case of a $Ge_{1-y}C_y$ alloy either random, diluted, or with C mostly arranged in third nearest neighbors. In the case of a pseudomorphic alloy on germanium (001), the tetragonal distortion of the lattice is well described (error $<0.13\%$) if the elastic constants of the alloy are linearly interpolated between the corresponding parameters of bulk Ge and C in the framework of macroscopic linear elastic theory. The position of the localized vibrational mode of substitutional C in Ge depends on the distribution of carbon, and new phonon modes appear for certain atomic arrangements. Carbon atoms in first nearest neighbor positions give new modes around 425, 480, and 610 cm^{-1}, while a new GeC-like mode occurs at 487 cm^{-1}. C in third nearest neighbor (most stable) arrangements induce an additional mode around 550 cm^{-1}. A signature of the fifth nearest neighbor configuration is calculated around 510 cm^{-1}. This study suggests that the content of substitutional carbon in $Ge_{1-y}C_y$ alloys can be estimated from the lattice constant by x-ray diffraction for $y<3\%$, and that the local order around C can be probed by Raman and infrared absorption spectroscopy. [S0163-1829(99)04046-1]

I. INTRODUCTION

Substitutional carbon incorporation into Si-Ge alloys seems to be promising for expanding the versatility of IV-IV semiconductor compounds and devices.[1] The Ge-Si-C alloy system is attractive for band gap and strain engineering of heterostructures.[2] Meanwhile, the lattice mismatch between C and Ge is very high (about 59%), the difference of electronegativity is important, and the solubility of C in Ge is very low (10^8 atoms/cm^3 at the melting point of Ge).[3] Therefore, careful optimization of the growth parameters is required to obtain a material suitable for device applications. Several attempts to grow $Ge_{1-y}C_y$ alloys have been reported,[1,4–7] but it is clear that carbon is not always fully substitutional in Ge. A theoretical analysis may help to predict the properties of the perfectly substitutional alloy, and several papers have addressed this issue.[8,9] The prior theoretical work, however, lacked the availability of recent experimental data.[5] In this present contribution, our theoretical approach will be explained; then the results about the structural effects of substitutional C in Ge and the corresponding localized phonon spectra will be successively detailed.

II. THEORETICAL APPROACH

Theoretical modeling of $Si_{1-x-y}Ge_xC_y$ alloys must take into account the strong local distortion around C. Kelires used a Tersoff potential[10] and Monte Carlo simulations[11] to predict significant deviation of the structural parameters from linearly interpolated values. In Kelires' approach, the interatomic potential was fitted to the calculated[12] enthalpy of formation ($\Delta H \approx 0.2$ eV/atom) of the hypothetical zinc-blende Ge-C alloy. The precision of this value might be questionable, and this approach does not yield the precise lattice parameter of carbon-free compounds. Alternatively, Rücker and Methfessel[13] fitted their anharmonic Keating potential to the lattice parameters, elastic constants, and zone-center optical phonon frequencies of the elementary compounds. This variant of the well-known Keating model[14] is adapted to the description of elastic properties, phonon frequencies, and relaxed geometry when large distortions are involved. In comparison, a full *ab initio* calculation would require greater computational effort. In our approach, we have kept the formalism and scaling laws of Rücker and Methfessel, but we have precisely fitted the force constants to the lattice parameters ($a_{Ge}=0.56576$ nm, $a_C=0.356683$ nm) (Ref. 15) and phonon modes of Ge[16] and diamond[17] [$\Gamma_{LTO}(\Gamma_{25'})_{Ge}=9.02$ THz, $\Gamma_{LTO}(\Gamma_{25'})_C=39.9$ THz]. The anharmonic Keating coefficients used are listed in Table I. The very high precision and reliability of these experimental data, which can be obtained by x-ray diffraction, Raman spectrometry, and absorption spectroscopy, justify our approach. The calculated lattice parameters and phonon modes are intended to have a precision of 10^{-5} nm and 0.1 cm^{-1}, respectively. Our potential has been also tuned to fit the recently measured[5,6] localized vibrational mode of C in Ge (531 cm^{-1}) and the predicted lattice parameter (a_{GeC}) of the hypothetical zinc-blende GeC compound. We previously obtained[18] a value of a_{GeC}

©1999 The American Physical Society

TABLE I. Anharmonic Keating parameters α and β used for the calculations. Force constants and lattice constants are given in atomic units. The values in parentheses [Rücker and Methfessel, Phys. Rev. B **53**, 3 (1996)] are given for comparison.

	a_0	α	β
C	6.7424 (6.68)	0.143 (0.137)	0.101 (0.101)
Ge	10.695 (10.66)	0.0555 (0.049)	0.015 (0.015)
GeC	8.54 (8.54)	0.0742 (0.081)	0.031 (0.031)

≈0.451 76 nm using a method similar to Ref. 19. The corresponding values found in the literature[19,9] are also close to 0.45 nm. To simulate the structure, molecular dynamics relaxation is calculated from a 512 atom cubic supercell by time increments of 3 fs, until the total potential energy reaches a stable minimum. To model the fully relaxed (i.e., virtually substrate-free) alloy, we have computed isotropic relaxations, where no external pressure is applied to the computational box. In addition, we have investigated pseudomorphic $Ge_{1-y}C_y$ alloys on Ge(001) to compare with recent experiments. By definition, the parallel lattice parameter of the supercell is maintained to the value of Ge here. In a further stage, the local phonon density around carbon is computed, using the recursion method detailed in Ref. 20. With standards notations, if the Green function for a given atom i and direction x is given by

$$G(\omega^2)=\left\langle i,x\left|\frac{1}{D-I\omega^2}\right|i,x\right\rangle, \tag{1}$$

then the local phonon density of state is defined by[20]

$$Y(\omega^2)=-\frac{1}{\pi}\lim_{\delta\to 0}\text{Im}[G(\omega^2+i\delta)]. \tag{2}$$

The recursion method permits to compute in a few steps a continuous fraction which approximates the Green function, thus enabling a convenient computation of the local phonon density of state for each atom. The calculated spectra can be compared with absorption spectroscopy and Raman spectrometry data after application of the proper selection rules. Several carbon distributions have been tested: random, diluted, and 3nn. By definition, the diluted distribution is obtained when the minimum distance between two carbon atoms is above 1 nm. The 3nn distribution corresponds to carbon atoms mostly in third nearest neighbor arrangements. This latter configuration is the most stable, in agreement with Refs. 13 and 9. The results presented here are typically averaged on 12 analogous distributions.

III. RESULTS AND DISCUSSION

A. Microstructure

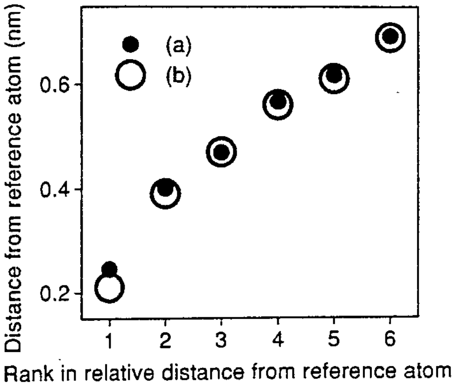


FIG. 1. Relative distance between nearest neigbors computed for (a) pure Ge, (b) one substitutional carbon atom surrounded by Ge in a cubic supercell of 512 atoms.

The extension of the local distortion around an individual C in the cubic supercell of 512 atoms is depicted in Fig. 1. The average distance between C and its first Ge neighbors is 0.213 62 nm, which is about 13% smaller than the Ge-Ge distance in pure Ge. The distortion is visible at the second nearest neighbor, but is attenuated after the third. Between the second and the third nearest neighbor, the Ge-Ge bond length increases to elastically accommodate the carbon-induced perturbation. It is clear that the carbon distortion is localized, and there is a local departure from a perfect translational symmetry, which induces a localized vibrational phonon mode studied below. This localized strain modifies the size of the supercell, and the average size along [001] gives the perpendicular lattice parameters represented in Fig. 2. In the case of isotropically relaxed alloys [Figs. 2(a) to 2(d)], the evolution of perpendicular (and parallel) lattice parameters follows Vegard's law (within 0.03%) up to 3% C, whatever the atomic distribution of C. This might appear surprising, since the lattice parameter of the reference zinc-blende GeC compound ($a_{GeC}\approx 0.451\,76$ nm) is smaller than the linear interpolation between Ge and C (≈0.461 22 nm). Actually, for the $Ge_{0.5}C_{0.5}$ alloy, we found that the statistical distribution of C has a strong influence on the lattice parameter. By comparison, the departure from Vegard's law is maximum around 50% Ge in Si-Ge alloy. The probability to form the SiGe ordered structure (with a smaller lattice parameter) may explain the measured deviation. For GeC alloys, Vegard's law does not apply to the

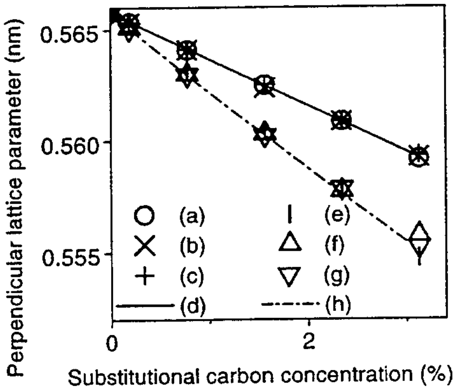


FIG. 2. Evolution of the perpendicular lattice parameter after incorporation of substitutional C in Ge. Several statistical distributions of C are considered: (a) C replacing Ge in a random alloy; (b) a diluted alloy, defined as when the interatomic distance between two carbon atoms is superior to 1 nm; (c) for carbon atoms mostly in third nearest neighbors (3 nn), the most stable (lowest energy) configuration; (d) the linear interpolation of the lattice parameters of Ge and diamond (Vegard's law). Curves (e)–(h) are analogous, except that the layer is pseudomorphic on Ge(001).

zinc-blende (1:1) compound, but this interpolation is reasonably accurate around Ge and diamond. Therefore it is possible to estimate accurately the substitutional carbon content of $Ge_{1-y}C_y$ alloys of low carbon content ($<3\%$) by measuring the lattice parameter with x-ray diffraction.[21,22] In the general case, the deviation from Vegard's law may depend on atomic distributions, especially around the regions of possible ordered compounds. This trend is exacerbated in the case of pseudomorphic layers on Ge (001). From Figs. 2(e) to 2(h), it is clear that the perpendicular lattice parameters (hence the elastic constants of the alloy) depend on the atomic distribution of atoms; therefore the atomic distribution of C modifies the relaxation process in the pseudomorphic lattice. Experimentally, the atomic distribution of atoms may depend on the growth process; therefore the elastic constants might also depend on the sample preparation. In all cases, a linear interpolation between the elastic constants and lattice parameters within the frame of macroscopic linear elastic theory [Fig. 2(h)] gives the perpendicular lattice parameter within a reasonable error ($<0.13\%$). This means that the relationship between the relaxed, parallel, and perpendicular lattice parameters (a_0, $a_{\parallel}$, $a_{\perp}$) may be expressed as

$$a_0=a_{\perp}\left(\frac{1+Ka_{\parallel}/a_{\perp}}{1+K}\right), \tag{3}$$

where

$$K=2\frac{\nu}{1-\nu}=2\frac{C_{12}}{C_{11}} \tag{4}$$

is the elastic correction factor and ν the Poisson ratio of the $Ge_{1-y}C_y$ alloy (with $y<3\%$). The departure from this law is maximum for the diluted alloy, where carbon atoms have nearly no interaction with each other. We conclude that the proximity of carbon atoms affects the global distortion of the pseudomorphic lattice. The cooperative effect of two close carbons seems to be superior to the sum of the effects of two isolated ones. The most probable (3nn) configuration gives a lattice parameter close to linear interpolation. The extrapolation to higher carbon concentrations increases the statistical spreading in lattice parameters and the subsequent deviations from Vegard's law, especially around the concentrations of possible ordered compounds.

B. Phonon spectra

The computation of phonon spectra is a further basic tool to investigate this new material. In Fig. 3, the localized vibrational spectra around C are computed in the case of an isotropically relaxed supercell containing two substitutional carbon atoms, for increasing C-C distance. In a way similar to $Si_{1-x}C_x$ alloys, novel modes appear for certain configurations. Localized vibrational spectra are sensitive to the local order and provide a powerful probe[8] of the bonding configuration. The 1nn configuration (improbable) gives new modes at 425, 480, and 610 cm^{-1}. GeC-like modes (2nn) are characterized by signatures at 487 and 540 cm^{-1}. The 3nn additional mode is around 550 cm^{-1}. Another additional feature, provided from the 5nn configuration, is situated around 510 cm^{-1}. Physically, these spectra can be considered as a result of complex conjugated elastic interactions involving close neighbors. The wave numbers depend slightly on the C content and the lattice strain. The shift of the main localized vibrational mode (LVM) of subtitutional C in Ge (around 530 cm^{-1}) depends upon the atomic distribution (Fig. 4). The local order may depend on the growth parameters, and there is no *a priori* shift of the mode with substitutional C content. In all cases, for low carbon contents, the LVM wave number of the pseudomorphic layer (around 529 cm^{-1}) is lower than its relaxed counterpart (around 531 cm^{-1}),[5,6] which is observed experimentally.

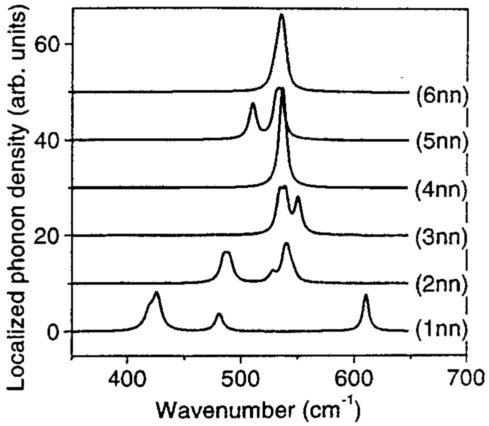


FIG. 3. Localized phonon density spectra around C calculated by the recursion method in the case of an isotropically relaxed supercell containing two carbon atoms. Curves (1nn)–(6nn) are calculated for increasing distance between C. With standard notation, "1nn" means that the two carbon atoms are arranged in first nearest neighbors. The width of the peaks can be numerically tuned; it is given for comparison purposes only.

IV. SUMMARY

In summary, we have performed molecular dynamics simulations using a valence force field model to simulate the incorporation of C into substitutional sites of Ge. In the case of relaxed alloys, the lattice parameter follows Vegard's law

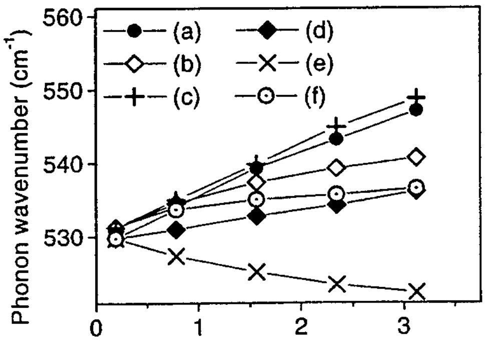


FIG. 4. Frequencies of localized phonon spectra around C calculated by the recursion method for increasing C content and various statistical distributions of substitutional C. Curves (a)–(c) are calculated for a random, diluted, and 3nn configuration, respectively, in the case of an isotropically relaxed alloy. The homologous curves (d)–(f) are computed for a pseudomorphic layer on Ge(001). The lines are displayed to guide the eyes only.

with a good accuracy (less than 0.03% error) up to at least 3% C, whatever the substitutional carbon distribution. Hence the carbon content can be successfully estimated through the measurement of lattice parameters by x-ray diffraction. The linear interpolation of elastic constants and lattice parameters of Ge and diamond gives a valuable estimate (error $<0.13\%$) of the perpendicular lattice parameter of the alloy within the framework of macroscopic linear elastic theory. The influence of the distribution statistics increases with C concentration, and a departure from Vegard's law is obtained around the region of possible ordered compounds (like zincblende GeC). The relative position of carbon neighbors influences the localized phonon spectrum, and the shift of the LVM wave number with C strongly depends on the carbon distribution. This study provides reference analytical formulas applying to the perfect $Ge_{1-y}C_y$ substitutional alloy. The predicted phonon spectra are a fundamental basis to probe the local order in this new material, which might play an important role in future IV-IV microelectronics and optoelectronics.

ACKNOWLEDGMENTS

We gratefully acknowledge support for this research by the ARO under Grant Nos. DAAH04-95-1-0625 and AASERT DAAG55-97-1-0249, by DARPA under Contract No. F49620-96-C-0006, and by the ONR under Grant No. N00014-93-1-0393.

---

*Author to whom correspondence should be addressed: CyrilGuedj@hotmail.com

[1] K. Brunner, K. Eberl, W. Winter, and E. Bugiel, Appl. Surf. Sci. **102**, 17 (1996).

[2] K. Eberl, S.S. Iyer, S. Zollner, J.C. Tsang, and F.K. Le Goues, Appl. Phys. Lett. **60**, 3033 (1992).

[3] R.I. Scace and G. Slack, J. Chem. Phys. **30**, 1551 (1959).

[4] H.J. Osten, E. Bugiel, and P. Zaumseil, J. Cryst. Growth **142**, 322 (1994).

[5] L. Hoffmann, J.C. Bach, B. Bech Nielsen, P. Leary, R. Jones, and S. Öberg, Phys. Rev. B **55**, 11 714 (1997).

[6] W.H. Weber, B.K. Yang, and M. Krishnamurthy, Appl. Phys. Lett. **73**, 626 (1998).

[7] J. Kolodzey, P.A. O'Neil, S. Zhang, B.A. Orner, K. Roe, K.M. Unruh, C.P. Swann, M.M. Waite, and S. Ismat Shah, Appl. Phys. Lett. **67**, 13 (1995).

[8] H. Rücker, M. Methfessel, B. Dietrich, K. Pressel, and H.J. Osten, Phys. Rev. B **53**, 1302 (1996).

[9] P.C. Kelires, Appl. Surf. Sci. **102**, 12 (1996).

[10] J. Tersoff, Phys. Rev. B **39**, 5566 (1989).

[11] P.C. Kelires, Phys. Rev. Lett. **75**, 1114 (1995).

[12] P.C. Kelires and P.J.H. Denteneer (unpublished).

[13] H. Rücker and M. Methfessel, Phys. Rev. B **52**, 11 059 (1995).

[14] P.N. Keating, Phys. Rev. **145**, 637 (1966).

[15] W. Kaiser and W.L. Bond, Phys. Rev. **115**, 857 (1959).

[16] G. Nilsson and G. Nelin, Phys. Rev. B **6**, 3777 (1972).

[17] S.A. Solin and A.K. Ramdas, Phys. Rev. B **1**, 1687 (1970).

[18] C. Guedj (unpublished).

[19] O.F. Sankey, A.A. Demkov, W.T. Petuskey, and P.F. McMillan, Modell. Simul. Mater. Sci. Eng. **1**, 741 (1993).

[20] A. Hairie, F. Hairie, G. Nouet, E. Paumier, and A.P. Sutton, in *Polycrystalline Semiconductors III—Physics and Technology*, edited by H.P. Strunk, J.H. Werner, B. Fortin, and O. Bonnaud (Trans Tech, Aedermannsdorf, Switzerland, 1993), Vol. 37-38, p. 91.

[21] J.P. Dismukes, L. Ekstrom, and R.J. Paff, J. Phys. Chem. **68**, 3021 (1964).

[22] R. Duschl, O.G. Schmidt, W. Winter, K. Eberl, M.W. Dashiell, J. Kolodzey, N.Y. Jin-Phillipp, and F. Phillipp, Appl. Phys. Lett. **74**, 1150 (1999).

Current Voltage Characteristics of High Current Density Silicon Esaki Diodes Grown by Molecular Beam Epitaxy and the Influence of Thermal Annealing

M.W. Dashiell, R.T. Troeger, S. L. Rommel, T.N. Adam, P. R. Berger and J. Kolodzey
University of Delaware, Department of Electrical and Computer Engineering, Newark, DE 19716 USA.

A. C. Seabaugh and R. Lake
Applied Research Laboratory, Raytheon Systems Company, Dallas, TX 75243 USA.

Abstract

We present the characteristics of uniformly doped silicon Esaki tunnel diodes grown by low temperature molecular beam epitaxy ($T_{growth} = 275$ °C) using *in situ* boron and phosphorus doping. The effects of *ex situ* thermal annealing are presented for temperatures between 640 and 800 °C. A maximum peak to valley current ratio of 1.47 was obtained at the optimum annealing temperature of 680 °C for 1 minute. Peak and valley (excess) currents decreased more than two orders of magnitude as annealing temperature and times were increased with rates empirically determined to have thermal activation energies of 2.2 and 2.4 eV respectively. The decrease in current density is attributed to broadening of the tunneling barrier due to the diffusion of phosphorus and boron. A peak current density of 47 kA/cm^2 was achieved and is the highest reported current density for a Si based Esaki diode (grown by either epitaxy or by alloying). The temperature dependence of the current voltage characteristics of a LTMBE Si Esaki diode in the range from 2.8 K to 325 K indicated that both the peak current and the excess current are dominated by quantum mechanical tunneling rather than by recombination. The temperature dependence of the peak and valley (excess) currents is due to the band gap dependence of the tunneling probability.

Submitted to IEEE Trans. Elec. Dev.

I. Introduction

To address issues of future device scaling and its limitations, researchers are focusing attention on new quantum devices, suitable for integration into silicon CMOS technology, with increased functionality, packing density and speed [1]-[5]. The negative differential resistance (NDR), bi-stability and high switching speeds associated with the quantum mechanical tunneling of electrons in devices including the Esaki tunnel diode, the resonant tunneling diode (RTD), and the resonant interband tunneling diode (RITD) may be exploitable for future logic, memory and oscillator circuits in computing and wireless communications applications.

Commercially available Si and Ge Esaki tunnel diodes are formed by alloying, a method incompatible with CMOS processing [5]. Few reports on a CMOS-compatible tunnel diode exist in the literature, such as Si RITDs grown by molecular beam epitaxy (MBE) [6,7], as well as p^+-i-n^+ diodes grown by MBE [8] and p^+-i-n^+ diodes combined with delta doping planes [9]. These diodes were grown at relatively low temperatures ranging from 325 to 370 °C. Two important figures of merits of tunnel diodes are the peak to valley current ratio (PVCR) and the peak current density. The highest PVCR for any epitaxially grown CMOS compatible tunnel diode is 4.2 (a Si/$Si_{0.5}Ge_{0.5}$ heterostructure Esaki with delta doping planes), while for an all Si diode it is 2.7 [9]. The addition of Ge to the intrinsic spacer was shown to increase the peak tunneling current as a result of lowering the tunneling barrier. The highest previously reported peak current density for a Si based tunnel diode was 22 kA/cm^2 for a Si/$Si_{0.5}Ge_{0.5}$ RITD grown at 370 °C. For an all Si RITD of equivalent structure and growth conditions, the peak current density was approximately one order of magnitude less [7]. A high current density is important for high speed switching applications where fast charging is necessary.

An obstacle to achieving heavily doped n^+/p^+ epitaxial layers by any technique is the low solid solubility of many electrical impurities (dopants) commonly used for Si. During heavy doping of MBE films, a profound effect is the thermally activated surface accumulation of impurities occurring at conventional Si growth temperatures, resulting in low dopant incorporation and spreading of the dopant profile [10], [11]. Recently, Gossman et al. reported on a low temperature MBE (LTMBE) ($T_{growth} < 300$ °C) technique which results in crystalline Si layers with 100% electrical activation of Sb and

B up to $6\times10^{20}cm^{-3}$ and $1\times10^{21}cm^{-3}$ respectively, for layers grown below a critical thickness, $h_{epi}(T)$ [12], [13].

There are disadvantages, however, associated with low temperature growth. Jorke et al. observed a significant increase in bulk recombination currents due to mid-gap states for Si diodes grown at 325 °C compared to diodes grown at 500 °C [8]. Several authors have reported high concentrations of point defects [14] or microvoids [15], which were detected using positron annihilation spectroscopy. For example, Gossman et al. [14] found a density of vacancy-like defects of $\sim10^{18}$ cm^{-3} for epitaxial Si films grown at 220 °C, while similar layers grown at 475 °C had a defect density of approximately 3 orders of magnitude less. However, these authors found that rapid thermal anneals (RTA) above 500 °C for 2 minutes reduced the defect density of the layers grown at 220 °C to below 5×10^{15} cm^{-3}, which is the sensitivity of the positron measurement.

In this paper, we present the properties of uniformly doped Si p^+-i-n^+ Esaki type diodes grown by LTMBE at 275 °C. We demonstrated a peak current density of 47 kA/cm^2 in a Si-only tunnel diode, which exceeds any previously reported Si based NDR device fabricated by epitaxy or alloying. We attribute our high current densities to the low growth temperature with high dopant incorporation. The PVCR of the diodes grown using this technique were slightly lower than reported in [9], possibly a result of defects associated with the extremely high doping concentrations in our diodes. We quantitatively discuss the influence of post-growth annealing on the current-voltage characteristics in the temperature range from 640 to 800 °C. The tunneling currents at constant biases were shown to decrease upon annealing with a rate characterized by thermal activation energies between 2.2 and 2.4 eV. Current- voltage measurements taken between 4.2 and 325 K reveal that the current is dominated by quantum mechanical tunneling for all applied voltage biases studied in this investigation.

II. Experimental

Uniformly doped Si Esaki diodes were grown by LTMBE in an EPI-620 MBE system. Si was evaporated from an electron gun at a rate of 0.5 nm/minute. P-type (boron) and n-type (phosphorus) dopants were evaporated from an elemental B source and a gettered compound GaP source [16] respectively. A 10 nm intrinsic Si spacer (i = nominally undoped) was grown between the n^+ and p^+ regions. The substrate temperature during growth was 275 °C, previously calibrated by observing the Au and Al eutetic reactions on a silicon substrate. A 15 nm p^+-type Si buffer ($N_A=1x10^{19}$ cm^{-3}) layer was grown on an 0.01 Ω-cm p-type (001) substrate, followed by the active regions: a 15 nm p^{++}-type Si layer, followed by a 10 nm i-layer, and finally the 30 nm n^{++}-type Si layer. Secondary Ion Mass Spectrometry (SIMS) revealed peak P concentrations of $6\text{-}8\times10^{20}$ cm^{-3} and B concentrations of 4×10^{20} cm^{-3} in the active regions of the diode. The i-layer thickness was calculated from the Si growth rate and does not take into account the possibility of bulk or surface dopant segregation.

The samples were annealed *ex situ* in an H_2/N_2 (15%/85%) ambient using a Heatpulse Rapid Thermal Annealing (RTA) furnace. Mesa diodes 15 μm in diameter were formed using standard photolithography and aluminum metallization, followed by junction isolation by etching in a CF_4/O_2 plasma. Room temperature electrical characteristics were measured using a HP4156B semiconductor parameter analyzer.

Samples for low temperature measurements were fabricated using a 40x40 μm^2 square mesa with Ti/Au metallization. Bond pads were formed on top of a polyimide-insulating layer and were subsequently wire bonded into a package compatible with the cryostat. The cryogenic chamber was a Janis Research Inc., model number 14-CNDT Cryostat and the temperature was controlled using a GaAs temperature sensor (1.4 - 330 °K) and a feedback controlled resistive heating element. The low temperature current voltage characteristics were measured with a HP4155A semiconductor parameter analyzer.

III. Results and Discussions

Current Voltage vs. Annealing

Figure 1 displays the SIMS dopant profile for the Esaki diode prior to *ex situ* annealings. The SIMS profile reveals a deviation from the step type dopant profile expected from the growth conditions. It is possible, however, that this smearing is an artifact of the SIMS measurement, since SIMS is known to suffer from knock-on and other effects which limit its depth resolution. Considering this uncertainty, it is difficult to quantify the affects of dopant segregation, which may have occurred during growth.

The inset of figure 1 displays the simulated band diagram near the junction under 0.1 V bias. *The band diagram was calculated by self-consistently solving the effective-mass Schrödinger equation and Poisson equation for the charge profile taken from figure 1 (it was assumed that the dopants were completely activated), but does not consider effects such as band-tailing or states within the bandgap.* As can be seen from the positions of the quasi-fermi levels relative to their respective band-edges, the doped layers are highly degenerate.

Figure 2 displays the room temperature current density vs. applied voltage characteristics (J-V) of the Si Esaki diode discussed above, after selected *ex situ* RTA anneals. The left-hand graph shows the J-V characteristics of the diode after 60 s isochronal RTA anneals. The right hand side shows the J-V characteristics after isothermal annealing at 775 °C. Negative differential resistance is observed for the lower three annealing temperatures (Figure 2, left) and for the shortest two anneal times (Figure 2, right). Due to the high current density, unannealed samples burned out before any NDR was observed, even under pulsed measurements. Table I lists the peak and valley current densities and their PVCR. Because of the high current densities the series resistance in the measurement setup (determined to be approximately 6 Ω) shifts the peak and valley currents to higher voltages. Analysis in the remainder of this section makes the appropriate correction for this voltage shift.

The diode currents in figure 2 are several orders of magnitude higher than expected from normal diode diffusion current. As a comparison, the left hand graph in figure 2 shows the simulated ideal diode current density for a Si p-n junction with $J_o = 1\mathrm{x}10^{-12}$ A/cm^2 (dotted line). Jorke observed deviations from ideal diode current in MBE grown p^+-i-n^+ Si diodes in the same bias range, for i-layer thicknesses of 40 nm and less [8]. They observed negative differential resistance for i-layers of 10 nm and 5 nm. These observations are consistent with earlier studies of alloyed junction Si Esaki diodes [17-19], where the current transport below the onset of NDR was shown to be due to phonon assisted tunneling transitions from conduction to valence band states (the band to band tunnel current). The current beyond the valley voltage, referred to as the excess current, was shown to be due to tunneling through defect levels in the forbidden energy gap up to a voltage where the thermal (i.e. diffusion or recombination) current begins to dominate. Both the band-to-band tunnel current and the excess current are proportional to an effective density of states and the probability for a tunneling transition to occur. The effective density of states for band-to-band tunneling can be expressed as an overlap integral between the density of states in the valence and conduction band multiplied by the occupational probabilities at a given energy [20]. Studies of the effect of high energy electron bombardment on the excess current on Si and Ge Esaki diodes revealed that the excess current increased proportionally to the electron dose. The chosen electron energy introduced defect energy levels within the material bandgap. Thus the magnitude of the excess current was proportional to the density of defect states within the gap [19]. The tunneling probability for both the band to band and excess tunneling currents are exponentially dependent on effective mass, the barrier height and width and field strength [20,21]. Expressions for the band to band tunneling current (J) and the excess tunneling current (J_x) are respectively:

$$J = Const \times D \times \exp\left\{ -\beta\, m^{*1/2} n^{*-1/2} E_G \right\} \tag{1}$$

$$J_x = Const \times D_x \times \exp\left\{ -\left(\alpha_x W_0 e^{1/2} \right)\left(E_G - eV + 0.6\left(V_n + V_p \right) \right) \right\} \tag{2}$$

The density of defect states is given by D_x, α_x is a material constant, e is the elementary charge, E_G is the band gap, V_n and V_p are the penetration of the quasi-fermi levels into the conduction and valence bands in volts, and V is the externally applied bias voltage across the junction. Equation 1 is a simplified expression for the band to band tunneling current using the abrupt junction approximation and assuming the junction potential is approximately equal to the bandgap voltage [21]. The parameters β and m^* (tunneling effective mass) are constants and D represents the density of states overlap integral at a given bias. Equation 2 is similar in form to equation 1 when the zero bias depletion width W_0 is substituted in terms of the effective carrier concentration n^* as in equation (3).

For the abrupt junction approximation, the zero bias depletion width is inversely proportional to the square root of the reduced or effective carrier concentration n^*, defined as:

$$\frac{1}{n*} = \left(\frac{1}{n} + \frac{1}{p}\right) \sim W_0^2 \qquad (3)$$

where n and p are the electron and hole concentrations on either side of the junction. Because $n^{*-1/2}$ is proportional to the depletion width, it is referred to as the junction width parameter. For alloyed junction Si Esaki diodes fabricated with various doping levels, both the band to band tunnel current at constant bias and the peak current resulted in a set of parallel lines when plotted semi-logarithmically against $n^{*-1/2}$ [18]. This is a consequence of the exponential dependence of the tunneling probability on the width of the tunneling barrier. In the case of alloyed junction Esaki diodes, the tunneling barrier is determined by the width of the depletion region.

We determined the effective values of n* versus annealing temperature for our high current density p^+-i-n^+ Si tunnel diodes by extrapolating the straight-line fit of $\ln J_{25mV}$ vs. $n^{*-1/2}$ in Figure 2 of reference 18 to our higher current densities (after correcting for series resistance). The bias of 25 mV is chosen to be consistent with Logan's data [18], and because at this bias the current transport is expected to be solely due to tunneling between conduction to valence band states; the contribution from the excess or defect current becomes significant only at higher biases. Note that the values of n* we have obtained by extrapolation from reference [18] do not necessarily reflect the actual electron and hole concentrations at any particular spatial position, since the depletion width of a p^+-i-n^+ structure will be both a function of the i-layer width and the carrier concentrations. Rather, we correlate the temperature dependence of our extrapolated n* values to an increase of the tunneling barrier width (using an abrupt junction approximation) after high temperature annealing.

Figure 3 depicts the natural logarithm of the current density at 25 mV (closed triangles) and 350 mV (open triangles) vs. $1/kT_{anneal}$ for 60 s duration *ex situ* annealing. The inset of figure 3 shows the decrease in current density at 25 mV and 350 mV vs. isothermal (775 °C) annealing time. An empirical expression for the tunneling current at constant bias is:

$$J = J_o \exp(-rt_{anneal}) \qquad (4)$$

with a thermally activated rate r:

$$r = r_o \exp\left(-E_A \big/ k_B T_{anneal} \right) \tag{5}$$

where J_o and r_o are constants, k_B is Boltzmann's constant and t_{anneal} and T_{anneal} are the annealing times and temperatures respectively.

Fitting the data in figure 3 to equations (4) and (5) (solid lines) gives values for the activation energy $E_A = 2.24 \pm 0.42$ eV for the current at 350 mV, and $E_A = 2.39 \pm 0.36$ eV for the current at 25 mV. As discussed previously, the current at 25 mV bias is assumed to be dominated by band-to-band tunneling. We assume also that at 350 mV the excess current is due to tunneling through defect levels as in reference [19]. The latter assumption is further validated in the following section where the current at a constant bias of 325 mV was shown by its temperature dependence to be a tunneling current. From the extracted activation energies and their uncertainties, it is apparent that both tunneling currents (band to band and defect related tunneling current) have nearly the same dependence on thermal annealing. This explains why the PVCR only varies by a factor of 1.5, while the absolute magnitudes of peak and valley currents vary by more than 2 orders of magnitude over the range of annealing temperatures. We note from equations (1) and (2) that the one common variable for both the band-to-band current and the excess current is the exponential dependence on the tunnel barrier width (via n^* in equation (1)). As discussed in more detail below, we conclude that for the annealing conditions used in this study, the change in magnitudes of the peak and valley currents are due to broadening of the tunnel barrier which occurs for anneals above 640 °C.

Duschl et al. has recently reported on the dependence on the peak and valley current on one minute anneals in the temperature range of 550 °C to 750 °C for heterostructure Si/$Si_{0.5}Ge_{0.5}$ p^+-i-n^+ Esaki diodes [9]. These authors observed two trends in the post annealing JV characteristics, where we distinguish between the low temperature (< 680 °C) and high temperature (> 680 °C) behavior. As the annealing temperature was increased from 550 to 680 °C the peak current of the Si/$Si_{0.5}Ge_{0.5}$ p^+-i-n^+ diode remained essentially constant, while the valley current decreased. In this temperature region, the PVCR increased from approximately 1 to the maximum value of

4.2 for a 680 °C anneal. These authors attributed this behavior to the annealing of electrical active point defects, which were formed during low temperature growth. A reduction in the number of these defects will decrease the density of defect states (D_X in equation (2)). Justification of their conclusion may be found in the study of Gossman et al. who observed that the density of vacancy-like defects in Si epitaxial layers grown by LTMBE was reduced by 3 orders of magnitude after annealing above 500 °C [14]. However no activation energy was reported for either study. Previous authors have correlated the annealing dependence of excess currents with a decrease in point defects in alloyed junction Si Esaki diodes. Logan et al. [23] examined the effect of electron induced damage to the lattice on the excess current. These authors observed the excess current at constant bias (350mV) was increased proportionally to the electron dose [19]. Post-bombardment annealing at temperatures from 300 to 400 °C reduced the excess current to its pre-bombardment magnitude at a rate characterized by an activation energy of 1.3 eV. The increase/decrease in excess current was attributed to the creation/annihilation of defect states within the bandgap (i.e. an increase/decrease of D_X in equation 2). During the electron irradiation/annealing experiments of reference [19] the peak current remained constant indicating that the width of the tunnel barrier had remained constant.

For annealing temperatures of 680 °C and above (the high temperature region) Duschl et al. found that the peak and valley currents decreased at the same rates and more rapidly than for the low temperature anneals. They attributed the decrease to a broadening of the depletion zone due to smearing of both the B and P delta doping spikes, which was confirmed by SIMS.

The high current density Si Esaki diodes that we investigated in this article were annealed under conditions similar to the high temperature anneal region in reference [9]. From the nearly equivalent thermal dependencies of the band-to-band and the excess currents, we concluded that the dominant mechanism occurring during *ex situ* annealing was broadening of the tunnel barrier width, which is consistent with Duschl's conclusions. For our uniformly doped p^+-i-n^+ Si Esaki diodes, we expect that the broadening of the tunneling barrier is due to dopant diffusion from the heavily doped p^+ and n^+ layers into the intrinsic spacer layer. Using the abrupt junction approximation, we calculated the relative changes in the tunnel barrier width in terms of the extrapolated n* (i.e. the width parameter) values listed in Table I. We then approximated a characteristic diffusion length of P and B into the intrinsic spacer assuming a Gaussian diffusion profile and using the diffusion coefficients for P and B in Si [22]. For all annealing conditions, the change in the tunnel barrier width determined from the extrapolated n* values were the same order of magnitude as calculated for the characteristic diffusion length of the P and B dopants into the intrinisic spacer layer. Based on the decrease in the 25 mV tunneling current from the value at 640 °C to that after 1 minute annealing at 680 °C (800 °C), the values of n* in table I correlated to an increase in barrier width of 2Å (25Å). Approximate calculations for the relative diffusion lengths for both P and B summed together are ~3 Å (20 Å) for one minute anneals at 680 °C (800 °C), in reasonable agreement with the observed decrease in current. The characteristic diffusion length calculation is approximate due to uncertainties in the diffusivities of P and B in the presence of excessive point defects and high concentrations.

We conclude then that for P and B doped Si Esaki diodes the dominant mechanism affecting the current voltage characteristics during annealing above 680 °C is dopant re-distribution, and the corresponding increase in the width, W_0, of the tunneling barrier. As dopants are depleted from the heavily doped p^+ and n^+ active regions, the spatial separation between the degenerate conduction and valence band edges increases, resulting in a strong decrease of both band-to-band (peak) and excess(valley) currents The change in magnitude of both currents exhibit nearly equivalent thermal dependencies. This is a direct consequence of the exponential dependence of the tunnel probability on the barrier width. We determined an empirical activation energy of between 2.2 and 2.4 eV to describe the rate for which the tunneling current magnitudes decrease due to the broadening of the P and B profiles in Si Esaki From reference [9], lower temperature anneals for one minute (≤ 640 °C) decrease only the valley current due to a reduction in point defects, thus increasing the PVCR. This process dominates only at low temperatures and may preliminarily be characterized by an activation energy of approximately 1.3 eV, based on the results of reference [23]. Above 680 °C, the redistribution of dopants appears to prevent further increases in PVCR since the peak current is dropping off at nearly the same rate as the valley current. The optimal anneal temperature for high PVCR in our study, as well as that in reference [9], is 680 °C. This temperature should correspond to that where defect annealing is most rapid, however; it is still low enough to prevent dopant redistribution from increasing the barrier width; thus the peak current remains essentially constant. The use of dopants with lower diffusivities, such as Sb and As, may permit higher temperature stabilty of the peak current and possibly greater PVCR since their diffusion coefficients are lower than for B and P in Si [22]. A low bulk diffusivity will allow *ex situ* anneals at higher temperatures and times so that defect annealing is maximized (excess current reduced), yet the tunnel barrier will remain unchanged (peak current remains constant). Silicon RITD structures synthesized using Sb and B delta doping planes indicate that Sb does not diffuse for one minute anneals at 700 °C, while the B significantly redistributes itself [24]. A significant difference for these Si RITDs is that the as grown diodes exhibited the IV characteristics of backward diodes. The peak current of these devices increased monotonically for annealing temperatures between 600 and 700 °C despite dopant redistribution. This is in direct contrast to their computer simulations, as well as to our results and those of reference [9] where as grown samples exhibited the highest current densities and *ex situ* annealing always decreased the currents. To explain the increase in peak current the authors suggested that annealing may be activating dopants in the delta doped planes [24].

Current Voltage Characteristics from 4.2 °K to 325 °K

The previous section suggested that both the peak and the excess currents were due to quantum mechanical tunneling. These assumptions were consistent with previous studies of alloyed junction Si Esaki diodes. It was deemed necessary, however, to confirm the transport mechanism(s) (for LTMBE Si Esaki's) since the LTMBE is a non-equilibrium process, while the alloying process occurs much closer to thermodynamic equilibrium. Since our growth process occurs far from equilibrium, the impurities and other defects may exist in concentrations significantly higher than by equilibrium growth methods and consequently may cause excessively high recombination currents. To verify the tunneling nature of the low bias (band to band) and high bias (excess) currents, we measured their dependence on diode operating temperature. The current voltage (IV) characteristics after annealing in forming gas 775 °C (60s) were measured as described in the experimental section. Figure 4 shows the current voltage curves of the Si Esaki diode with $i \approx 10$ nm diode at measurement temperatures from 4.2 °K to 325 °K. As the temperature increased, a variation of less than one order of magnitude in current was observed. A series resistance of 8 Ω in this measurement setup accounts for the increase in peak voltage as the peak current is increased. A weak temperature dependence of the current is typical for tunneling, which is not thermally activated in contrast to the normal diode diffusion or recombination current. Figure 5 plots the PVCR versus measurement temperature (left axis) and the ratios of the peak, valley and constant bias (325 mV) currents to their respective 4.2 °K current magnitudes. It is evident that the peak and valley currents exhibit almost exactly the same temperature dependence up to about room temperature. The excess current at constant bias (325 mV-corrected for series resistance) likewise exhibits a weak temperature dependence.

The ratio of the peak tunneling current to its value at 4.2 °K can be derived from equation (1) in the previous section

$$\frac{I_{tunnel}(T)}{I_{tunnel}(4.2^{o}K)} = \exp\left[-\beta\, m^{*\frac{1}{2}} n^{*-\frac{1}{2}} \Delta E_G\right] \qquad (6)$$

where ΔE_G is the bandgap difference between temperature T and its value at 4.2 K, m^* and n^* are the effective electron mass and reduced carrier concentration, and β is a constant. The exponent of 1 for the bandgap term is again due to the approximation of the built in voltage being equal to that of the band gap. We have fit the temperature dependence of the peak current ratios to the above equation using the optical bandgap temperature dependence data of [25] to obtain the solid line shown in figure 5. A value of 5.1 eV^{-1} is obtained for $\beta m^{*1/2} n^{*-1/2}$. The good fit between the experimental curves of $I_{peak}(T)/I_{peak}(4.2\ °K)$ and equation 6 illustrate that the temperature dependence of the tunneling currents expected from the tunneling transition probability, where the change in bandgap is the only temperature dependent term. The slight deviation (decrease) of the experimental points from theory near room temperature may indicate that the overlap integral between the fermi-dirac occupational probability and the density of available states is changing due to the temperature dependence of the fermi-level. It is well known that as temperature increases, the fermi level decreases towards the intrinsic fermi level until eventually the material becomes non-degenerate. The good agreement between the tunneling probability term and the experimental data at temperatures below ~300 K suggests that the overlap integral remains constant up to approximately room temperature.

The valley and the excess currents at a constant bias of 325 mV (corrected for series resistance) likewise follow an exponential dependence on Si bandgap up to around room temperature, implying the tunneling nature of the excess current. In figure 5, the deviation from the theoretical curve from the temperature dependence of the excess current ratio may be explained by the fact that the factor β in in equation (6) actually depends on bias through the field strength at the junction. Thus over the temperature and bias conditions we have studied, we rule out recombination currents inside the depletion region or at the surface as contributors to the excess currents because these currents exhibit strong thermal dependencies.

To observe the fine structure of the IV characteristics, we have submerged the sample in liquid helium and evacuated the submersion chamber to achieve a measurement temperature of 1.7 °K. An observable increase in the conductance at energies corresponding to certain phonon (or combination of phonons) energies is expected in an indirect material such as Si due to momentum conservation [17]. Figure 6 shows the first and second derivatives of the IV characteristics at biases from -200 mV to +300 mV. Three inflections are evident in the first derivative curve and become more pronounced in the second derivative curve where the phonon energy is approximately indicated by the maximum of the second derivative peak [17]. For the diode shown in figure 4, these values occur at -66, +53 and +173 mV. A correction of 8 Ω for the series resistance give values of -17, +18 and +49 mV. These values are reasonably close to the values assigned by Chynoweth for the transverse acoustic (18.4meV) and the transverse optical (57.6 meV) phonon branches. The difference between Chynoweth's TO value and ours may be due to incomplete voltage correction caused by a voltage-dependent series resistance for our non-alloyed contacts. It is probable that at 1.7 °K, a non-alloyed contact exhibits some non-linearity due to thermionic emission through the metal-semiconductor barrier. Other than the precise positions of the inflection points in the first derivative of the IV, our conductance curves appear very similar to those of Chynoweth's. These observations indicate that the low bias transport mechanisms are the same for LTMBE Esaki diodes and for alloyed junction diodes, despite our much higher current densities and low temperature growth process.

IV. Conclusions

We have presented current-voltage results and the effects of *ex situ* rapid thermal annealing on uniformly doped Si Esaki diodes grown by MBE. We have shown that all tunneling currents are reduced by *ex situ* annealing for one minute between 640 and 800 °C. The decrease in tunneling currents is attributed to an increase of the tunnel barrier due to P and B diffusion in Si. The rate of current decrease is described by a thermal activated rate with an activation energy of between 2.2 eV and 2.4 eV. An optimal annealing temperature of approximately 680 °C for maximal PVCR in Si Esaki diodes doped with P and B appears to occur at the transition temperature where the mechanism dominating the *ex situ* IV characteristics changes from defect annealing to dopant redistribution. This behavior is expected when comparing our activation energy to the lower activation energy (1.3 eV) describing the annealing of point defects in Si Esaki diodes [23]. These conclusions suggest that higher PVCR are obtainable by using dopants with lower diffusivities such as As and Sb, since higher temperatures may be used for point defect annealing, while maintaining the original doping profile.

Variable temperature IV measurements were used to verify that both the peak current and excess currents were dominated by tunneling currents rather than thermally activated recombination or diffusion currents. Inflections in the diode conductance were observed at cyrogenic temperatures, consistent with phonon assisted tunneling in an indirect band gap semiconductor. The measured peak current density of 47 kA/cm^2 is higher than that reported for any Si based tunnel diode including epitaxial Si/SiGe tunnel diodes and alloyed junction Si Esaki's. Only one letter reports a higher peak current density (~100kA/cm^2) in a Si Esaki, by combining a heavy diffusion process with an alloying technique, however their actually current density may be significantly lower due to the large uncertainty and non-uniformities in determining the junction area.[26].

Acknowledgements

The authors gratefully acknowledge support by DARPA contract number F49620-96-C-0006 and by ONR grant no. N00014-93-10393. They thank R. G. Wilson for SIMS measurements, C. Guedj for x-ray diffraction measurements, M. Green for processing assistance and P. Thompson and Karl Hobart for discussions concerning MBE growth calibration. They thank Bob Aldret for invaluable assistance with the low temperature measurements. They also thank Mike Mauk, Paul Sims, Brian Feyock and Jeff Cox of AstroPower Inc. for assistance with plasma etching and the polyimide process.

References

[1] A. Seabaugh, R. Lake, B. Brar, R. Wallace and G. Wilk, “Beyond the roadmap technology: Silicon heterojunctions, optoelectronics and quantum devices”, Mat. Res. Soc. Symp. Proc., vol. 486, pp. 67-77, 1998.

[2] J.P.A. Vanderwagt, "Tunneling Based SRAM", Proceedings of the IEEE, vol 87, pp. 571-595, 1999.

[3] R.H. Mathews, J.P. Sage, T.C.L. Gerhard Sollner, S.D. Calawa, C.L. Chen, L.J. Mahoney, P.A. Maki, K.M. Molvar, "A new RTD FET Logic Family", Proc. IEEE, vol. 87, pp. 596-605, 1999.

[4] P. Mazumber, S. Kulkarni, M. Bhattacharya, J.P. Sun and G.I. Haddad, “Digital Circuit Applications of Resonant Tunneling Devices”, Proc. IEEE, vol. 86, pp. 664-686, 1998.

[5] A. Seabaugh and R. Lake, “Tunnel diodes”. Encyclopedia of Applied Physics, vol. 22, pp. 335-358, 1998.

[6] S.L. Rommel, T.E. Dillon, M.W. Dashiell, H. Feng, J. Kolodzey, P.R. Berger, P.E. Thompson, K.D. Hobart, R. Lake, A.C. Seabaugh, G. Klimeck, and D.K. Blanks, “Room temperature operation of epitaxially grown Si/SiGe/Si resonant interband tunneling diodes”, Appl. Phys. Lett., vol 73, pp. 2191-2193, 1998.

[7] S.L. Rommel, T.E. Dillon, P.R. Berger, R. Lake, P.E. Thompson, K.D. Hobart, A. Seabaugh and D. Simons, “Si-Based Interband Tunneling Devices For High-Speed Logic and Low Power Memory Applications”, International Electron Devices Meeting Technical Digest, pp. 1035-1037, 1998.

[8] H. Jorke, H. Kibbel, K. Strohm, E. Kasper, “Forward-bias characteristics of Si bipolar junctions grown by molecular beam epitaxy at low temperatures” Appl. Phys. Lett., vol. 63, pp. 2408-2410, 1993.

[9] R. Duschl, O.G. Schmidt, G. Reitemann, E. Kasper and K. Eberl, “High room temperature peak-to-valley current ratio in Si based Esaki diodes” Elect. Letters, vol. 35, pp. 1111-1112, 1999.

[10] C.P. Parker, R.A. Kubiak, S.M. Newstead, T.E. Whall and E.H.C. Parker, “Temperature dependence of incorporation processed during heavy boron doping in silicon molecular beam epitaxy”, J. Appl. Phys., vol. 71, pp. 118-125, 1992.

[11] E. Friess, J. Nutzel, and G. Abstreiter, “Phosphorus doping in low temperature silicon molecular beam epitaxy”, Appl. Phys. Lett., vol. 60, pp. 2237-2239, 1992.

[12] H.J. Gossman, F.C. Unterwald and H.S. Luftman, “Doping of Si thin films by low-temperature molecular beam epitaxy” J. Appl. Phys., vol 73, pp. 8237-8241, 1993.

[13] D.J. Eaglesham, H.J. Gossman, and M. Cerullo, “Limiting thickness h_{epi} for epitaxial growth and room-temperature Si growth on Si (100)”, Phys. Rev. Lett., vol. 65, pp. 1227-1230, 1990.

[14] H.-J. Gossman, P. Asoka-Kumar, T.C. Leung, B. Nielsen, K.G. Lynn, F.C. Unterwald and L.C. Feldman, “Point defects in Si thin films grown by molecular beam epitaxy”, Appl. Phys. Lett., vol. 61, pp. 540-542, 1992.

[15] D.D. Perovic, G.C. Weatherly, P.J. Simpson, P.J. Schultz, T.E. Jackman, G.C. Aers, J.-P. Noel and D.C. Houghton “Microvoid formation in low-temperature molecular beam epitaxy grown Si”, Phys. Rev. B, vol. 43, pp. 14257-14260, 1991.

[16] G. Lippert, H.J. Osten, D. Kruger, and P. Gaworzewski, “Heavy phosphorus doping in molecular beam epitaxial grown silicon with a GaP decomposition source”, Appl. Phys. Lett., vol. 66, pp. 3197-3199, 1995.

[17] A.G. Chynoweth, R.A. Logan, and D.E. Thomas, “Phonon-assisted tunneling in silicon and germanium esaki junctions”, Phys. Rev., vol. 125, pp.877-881, 1962.

[18] R.A. Logan and A.G. Chynoweth, “Effect of degenerate semiconductor band structure on current-voltage characteristics of silicon tunnel diodes”, Phys. Rev., vol. 131, pp. 89-95, 1963.

[19] A.G. Chynoweth, W.L. Fellman and R.A. Logan, “Excess tunnel current in silicon Esaki junctions”, Phys. Rev., vol. 121, pp. 684-713. 1961.

[20] S. Sze in “Physics of Semiconductors” 2nd ed., Wiley, New York, 1985.

[21] D. Meyerhoffer, G.A. Brown, H.S. Sommers, "Degenerate germanium I. Tunnel, excess and thermal current in tunnel diodes", Phys. Rev. vol. 26, pp. 1329-1341, 1962.

[22] S.M. Sze, "VLSI Technology" Mc-Graw-Hill, New York, p. 286, 1988.

[23] R.A. Logan, W.M. Augustyniak and J.F. Gilbert, “Electron bombardment damage in silicon Esaki diodes”, J. Appl. Phys., vol. 32, pp. 1201-1205. 1961.

[24] P.E. Thompson, K.D. Hobart, M.E. Twigg, G. G. Jernigan, T.E. Dillon, S.L. Rommel, P.R. Berger, D.S. Simons, P.H. Chi, R. Lake, A.C. Seabaugh, “Si resonant interband tunnel diodes grown by low-temperature molecular beam epitaxy”, Appl. Phys. Lett., vol. 75, pp. 1308-1310, 1999.

[25] G.G. Macfarlane, T.P. McLean, J.E. Quarrington, V. Roberts “Fine structure in the absorption edge spectrum of Si” Phys. Rev., vol. 111, pp. 1245-1254 (1958).
[26] H. Hornung and D. Zook “High speed diffused-alloyed silicon tunnel diodes”, Solid-State Electronics, vol. 9, pp. 7-12, 1966.

Figure and Table Captions

Figure 1. SIMS profile of B and P dopants for the high current density Si p^+-i-n^+ Esaki diode prior to *ex situ* annealing. The inset displays the calculated band diagram where E_F is the quasi fermi level, E_C is the conduction band and E_V is the valence band.

Figure 2. Current density vs. the applied voltage for a LTMBE Si tunnel diode after 60 s anneals at various temperatures (left hand plot). The right hand plot displays the same diode structure after annealing at 775 °C for various times. The dotted line simulates the magnitude of the ideal diffusion diode current in Si for $J_o = 1\times10^{-12}$ A/cm^2.

Figure 3. Plot of the natural logarithm of current density at constant biases vs. the inverse of the annealing temperature (T_{ANN}) where k_B is the Boltzmann factor. The open triangles represent the current at 350 mV, while the filled triangles represents the current at 25 mV after series resistance corrections. The data was fit to equations (4) and (5) giving an activation energy of 2.24 eV ± 0.42 eV and 2.39 ± 0.36 eV. The inset to figure 2 shows the decrease in current as a function of isothermal annealing time.

Figure 4. Variable temperature IV curves of a 40×40 μm^2 Si Esaki tunnel diode after a 60 s anneal at 775 °C. Measurement temperatures ranged from 4.2 °K to 325 °K. The variation of peak voltage versus peak current yields a series resistance of 8 Ω.

Figure 5. Peak-to-valley current ratio (PVCR) of the diode from figure 4 as a function of measurement temperature (filled circles, left axis). The right axis shows the correct scale for the ratio of the peak (hollow circles) and valley (hollow up-triangles) currents as well as the current at a constant bias of 325 mV (hollow squares) to their respective current values at 4.2 °K. The solid line is a fit to equation 6 using the optical bandgap data of reference [24], illustrating that the temperature dependence of the tunneling current is due to the change in bandgap in the tunneling probability. The slight deviation of the valley and high bias (350 mV) points from the theoretical curve is expected at higher voltages, due to increases in field strength in the junction.

Figure 6. Conductance, dI/dV (left axis) and the second derivative, d^2I/dV^2 (right axis) of the LTMBE grown Si Esaki diode from figure 4 taken at 1.7 °K. After correcting for series resistance, we attribute the inflections to contributions from the TA (18.4 meV) and TO (57.6 meV) phonons to the tunneling current.

Table I. Relevant dc electrical parameters for MBE grown Si Esaki diode structure A for one minute RTA anneals. The values of the reduced carrier concentration, n^* were extrapolated from the data given in reference [18]. *The diode annealed at 800 °C did not exhibit a valley region, only an inflection point.

Table I.

Anneal Temperature °C	J_{peak} (kA/cm^2)	J_{valley} (kA/cm^2)	PVCR	Extrapolated n^* ($\times 10^{19}$ cm^{-3})
640	46.8 ± 3.1	35.8 ± 1.0	1.3 ± 0.06	4.9
680	43.4 ± 1.9	29.2 ± 1.3	1.47 ± 0.05	4.8
700	26.1 ± 0.7	18.4 ± 1.3	1.425 ± 0.06	4.4
720	19.0 ± 1.9	14.1 ± 1.2	1.34 ± 0.04	4.0
750	9.3 ± 0.2	7.9 ± 0.09	1.17 ± 0.01	3.7
775	3.6 ± .05	3.2 ± 0.06	1.1 ± 0.01	3.5
800	0.05*	-	-	2.2

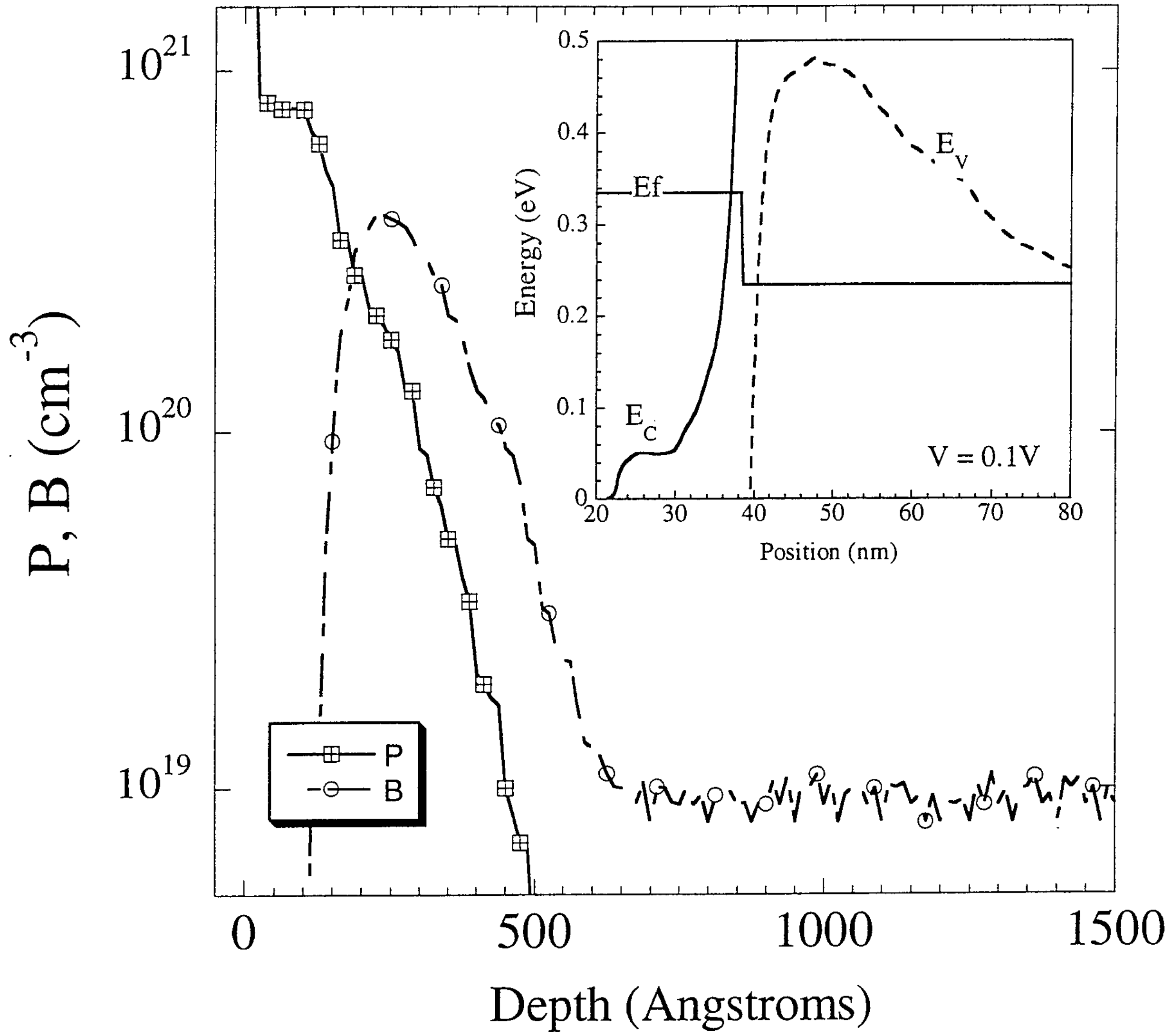

10^{21}
10^{20}
10^{19}
P, B (cm^{-3})
0
500
1000
1500
Depth (Angstroms)
P
B
0.5
0.4
0.3
0.2
0.1
0
Energy (eV)
20
30
40
50
60
70
80
Position (nm)
Ef
E_C
E_V
V = 0.1V

Figure 1

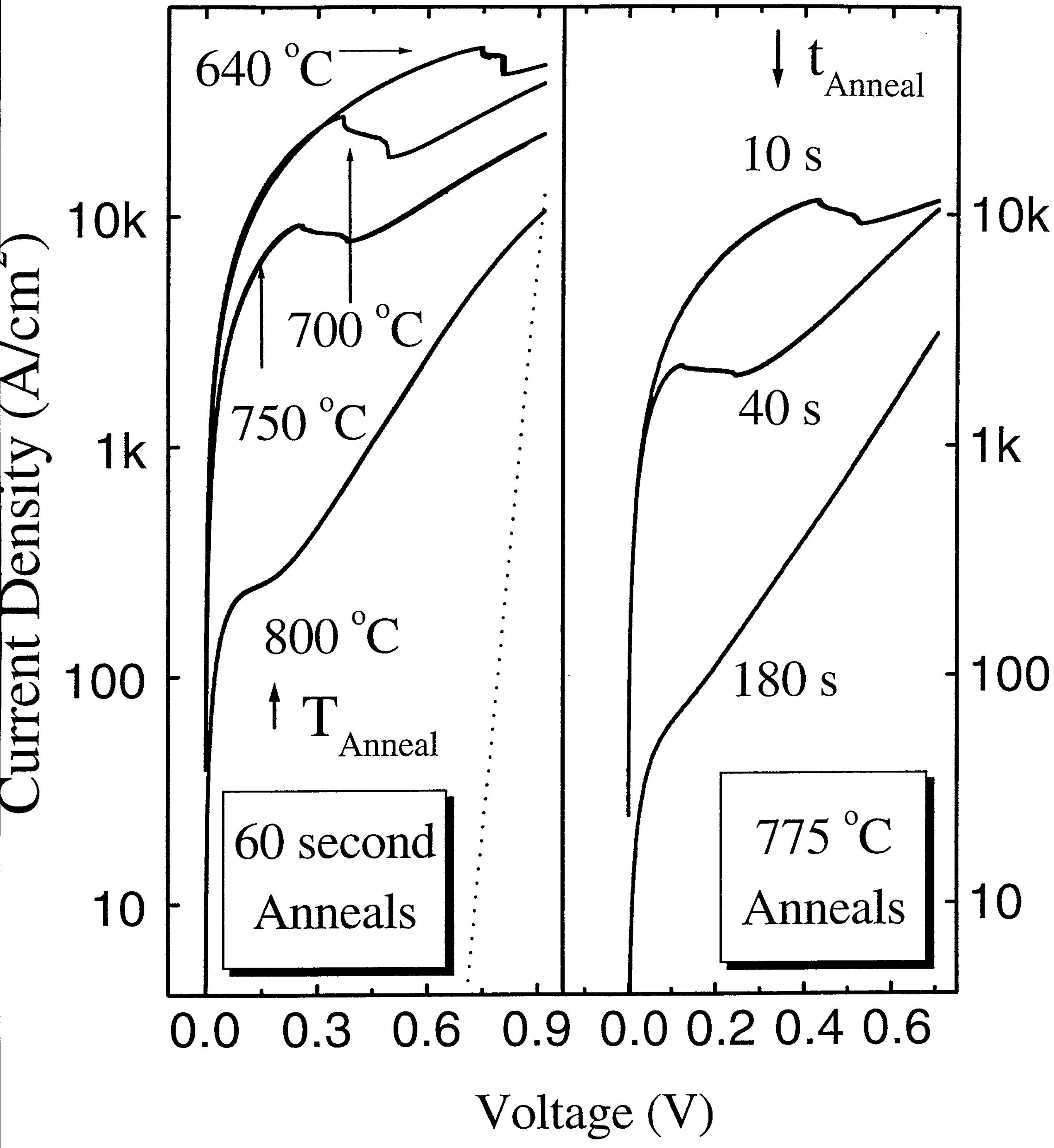

640 °C
700 °C
750 °C
800 °C
↑ T_{Anneal}
60 second Anneals
↓ t_{Anneal}
10 s
40 s
180 s
775 °C Anneals
Current Density (A/cm^2)
10k
1k
100
10
0.0
0.3
0.6
0.9
0.0
0.2
0.4
0.6
Voltage (V)

Figure 2

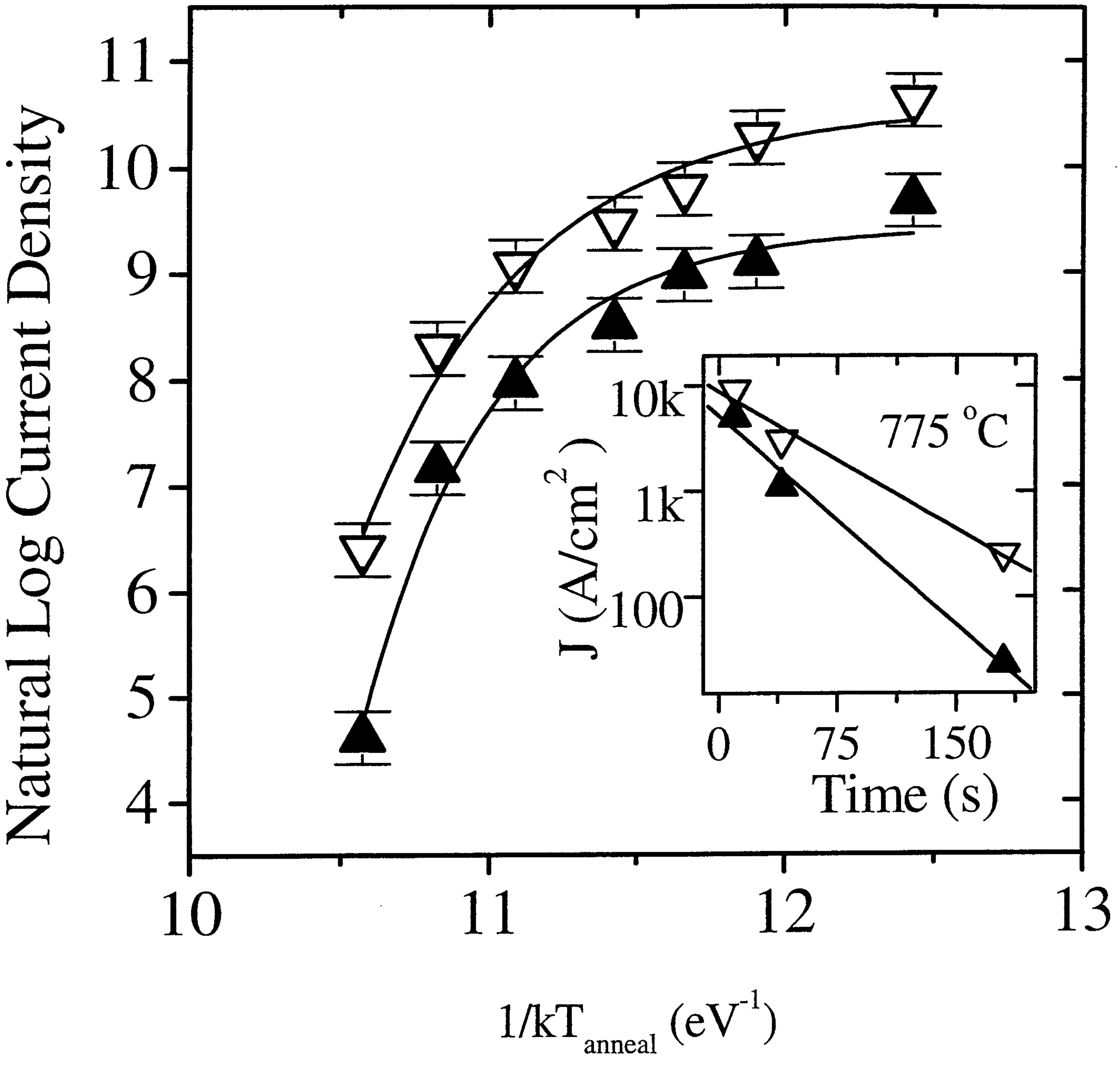

Natural Log Current Density
11
10
9
8
7
6
5
4
10
11
12
13
$1/kT_{anneal}$ (eV^{-1})
J (A/cm^2)
10k
1k
100
0
75
150
Time (s)
775 °C

Figure 3

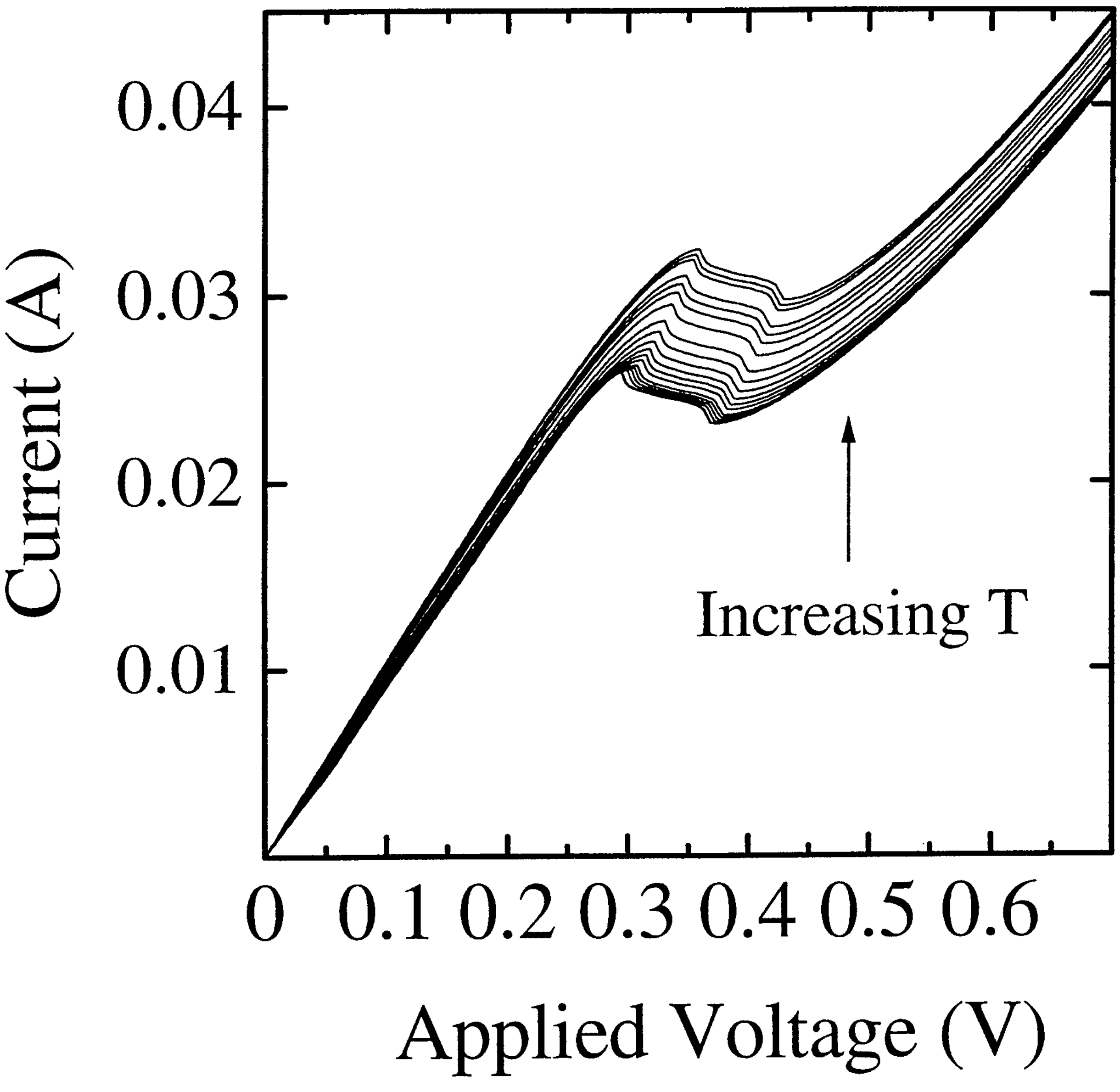

Current (A)
0.04
0.03
0.02
0.01
0
0.1
0.2
0.3
0.4
0.5
0.6
Applied Voltage (V)
Increasing T

Figure 4

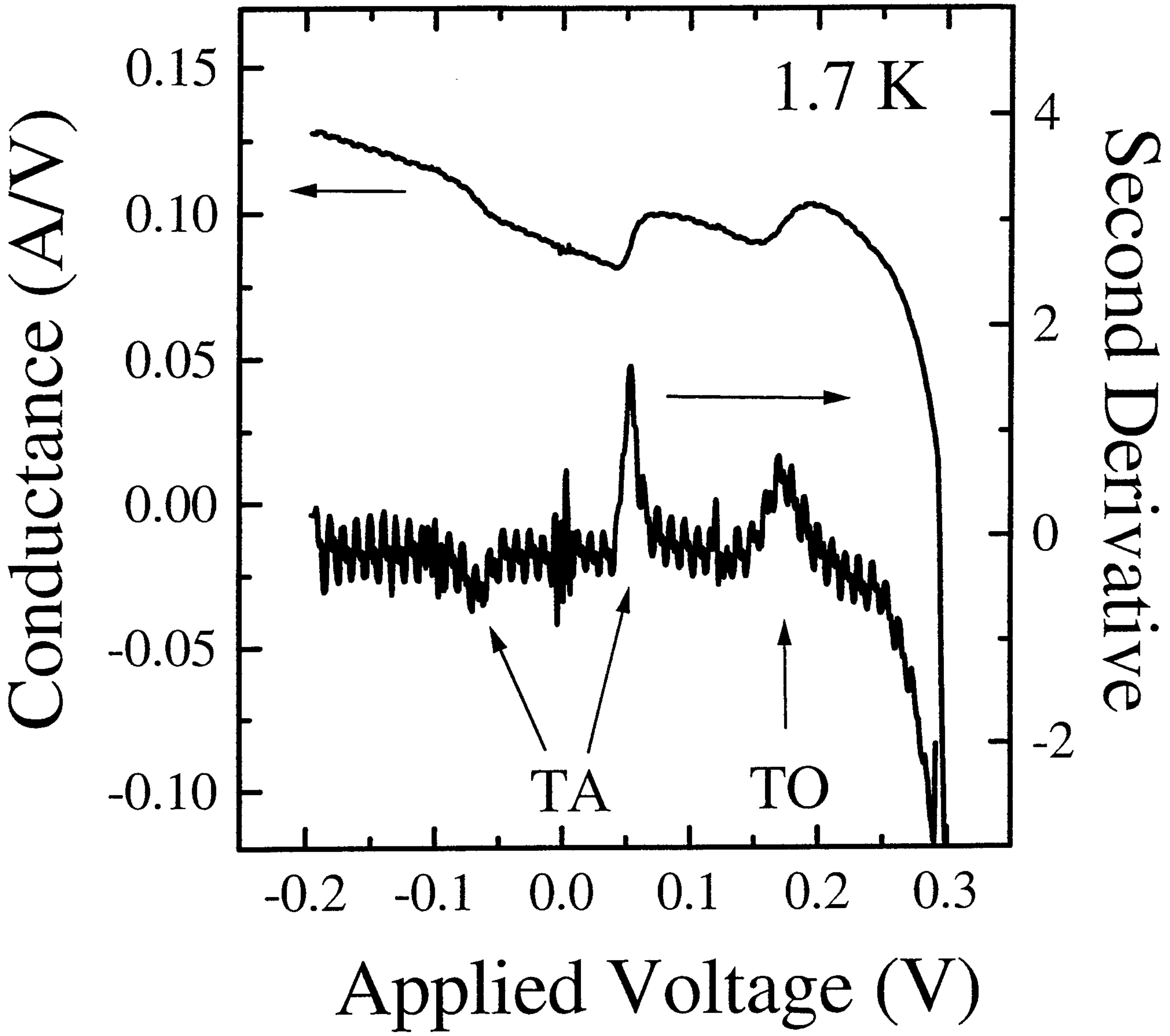

1.7 K
Conductance (A/V)
0.15
0.10
0.05
0.00
-0.05
-0.10
Second Derivative
4
2
0
-2
TA
TO
-0.2
-0.1
0.0
0.1
0.2
0.3
Applied Voltage (V)

Figure 6

Circuit Applications of Quantum MOS Logic*

Shriram KULKARNI and Pinaki MAZUMDER†

Abstract

We present design approaches and system applications for circuits that integrate resonant tunneling diodes (RTDs) with MOS devices in the post-shrinking VLSI era to achieve improved performance as compared to conventional CMOS. The high switching speed of RTDs and their negative differential-resistance (NDR) characteristics are ideally suited for the design of fast and compact self-latching logic circuits. The possible use, in the future, of these quantum MOS (QMOS) gates in deeply pipelined computing systems will result in elimination of the area, delay and power overhead of pipeline latches that limits conventional CMOS-based systems, leading to compact and fast system designs while retaining advantages of CMOS, such as low power dissipation and high packing density. In particular, communication systems and digital signal processors are expected to benefit from the gate-level pipelining approach due to their low data dependence.

1 Introduction

State-of-the-art technologies based on field-effect devices such as MOSFETs have, over the past several decades, met growing end user performance requirements by scaling device dimensions. The physical limits on performance improvement by scaling, using conventional device transport phenomena, will likely be reached in the early part of the next century, necessitating alternative device concepts to continue fueling the growth of the VLSI industry. As device dimensions in integrated circuits shrink to deep-submicron levels, quantum effects, such as resonant tunneling, become more prominent leading to interesting new device characteristics which can be exploited to create extremely fast and compact circuits [1]. Several novel logic circuits using RTDs in conjunction with transistors [2, 3, 4, 5] have been proposed in the past few years. These circuits have one or more of the following advantages over conventional logic circuits: 1) reduced circuit complexity for implementing a given function, 2) low power operation, and 3) high speed operation. RTD-based circuits are usually fabricated using III-V technology which offers high performance devices such as hetero-junction bipolar transistors (HBTs) and high electron mobility transistors (HEMTs). While such circuits demonstrate very high switching speeds, they operate at high current levels thus consuming substantially higher *absolute* power than equivalent CMOS circuits. Also, the packing density of III-V technology based VLSI circuits is more than an order of magnitude smaller than CMOS VLSI circuits. However, the unique NDR characteristics of RTDs coupled with their high tunneling speeds lead to very compact and fast circuit topologies. Thus it is very attractive to envision these compact, high-functionality circuits implemented in a technology such as CMOS that offers low power dissipation and very large integration levels. While co-integration of RTDs and MOSFETs is still an active area of research, in the following sections, we discuss the prospects of circuits and systems using QMOS technology using specific examples and simulation results.

2 QMOS Logic Families

Two basic QMOS logic families are: 1) Static QMOS, and 2) Bistable QMOS. A static QMOS gate consists of an RTD pull-up load and a pulldown network of n-transistors that determines the logic operation performed by the gate. All static QMOS gates may be designed as ratioless circuits. This feature is of use in certain layout design styles such as sea-of-gates and gate arrays. Static QMOS logic yields the simplest and most robust logic circuits of any of the QMOS logic family. The circuit output high voltage is equal to the power supply voltage, V_{dd}, whereas the output low voltage is determined by the device characteristics of the RTD and the FET. Static QMOS circuits rely solely on the fast switching of the RTD across its NDR region for performance improvement. The possibility of two stable RTD operating points at the same current level, due to NDR characteristics, makes bistable QMOS logic feasible. The operating principle of the bistable element may be understood by considering the simplified circuit shown in Figure 1(a). An RTD forms the active load in the circuit. There are *m* inputs in the n-type MOSFET logic block which determines the circuit function, a clock transistor which controls the evaluation of the gate, and a bias transistor that maintains the quiescent current through the RTD while also controlling the precharging of the gate output. Figure 1(b) shows the load lines for the bistable logic gate. The bias transistor is used to reset the state of the gate or to maintain a quiescent current through the RTD which is between its peak and valley currents. Initially, a low going pulse on the bias line resets the gate output to logic 1 (point

---

*This research has been supported by a grant from DARPA.

†Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122 USA

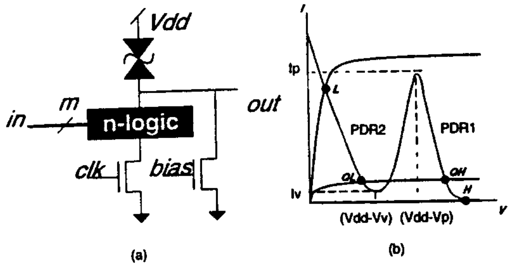


Figure 1: (a) Bistable-mode QMOS logic gate (b) Load lines for bistable QMOS gate operating point

H in the load lines). The clock transistor is designed such that if the n-block is turned on, turning the clock transistor on will cause the current through the RTD to exceed its peak current value thus switching the state of the gate and causing the output to go low. Thus, any change in the inputs $I_1, I_2, ...I_m$ is reflected at the output node only when the clock signal is active. When the clock signal becomes inactive, the bias transistor maintains the output voltage within the correct range of logic 1 or logic 0. The circuit operating point is either QH or QL depending on whether the n-logic block is off or on when the clock signal becomes inactive. Figure 2 shows the simulation of a QMOS bistable inverter circuit operating at a clock frequency of 10 GHz. The simulation of QMOS circuits has been carried out using a SPICE simulator for NDR devices [6]. The circuit has been designed using 0.35 micron CMOS devices and operates at a power supply voltage of 1.5 V with a power dissipation of 70.7 μW. The absence of the threshold drop of the PMOS devices allows for lower supply voltage operation in QMOS circuits. The power-delay product of 7.07 fJ is much superior to a corresponding dynamic CMOS circuit with an identical pulldown network that has a power-delay product of 10 fJ operating at a maximum clock frequency of 5 GHz. Additionally, the RTD maintains the output node voltage via a conducting path from the supply and hence has greater noise immunity that dynamic CMOS circuits that are susceptible to charge sharing errors.

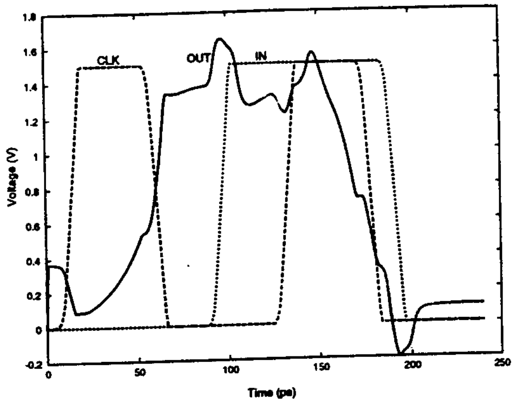


Figure 2: Simulation of bistable QMOS inverter at 10 GHz clock

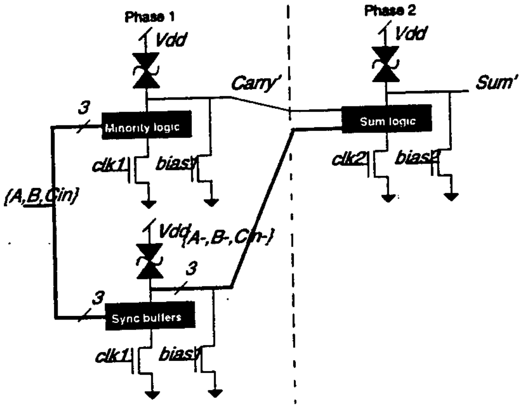


Figure 3: Pipelined RTD-MOSFET full adder schematic

3 Gate-level Pipelining of QMOS Circuits

Using QMOS logic, a bistable full adder is designed as a two-stage logic block. For correct operation of the bistable logic gates, bias and clock pulses are required as mentioned previously. However, when multiple gates are cascaded, as is the case in a full adder, a gate must be clocked only after all its inputs have been correctly evaluated. This requires a multiphase clocking scheme in which each gate is evaluated in a different phase than its fanins and fanouts. For cascaded RTD-MOSFET gates, a two-phase clocking scheme is used. This is similar to dynamic CMOS clocking. Figure 3 shows the schematic of the full adder circuit. The majority function that gives the carry output is evaluated on $Phase1$ of the clock. Each clock phase consists of a pulse on the bias line followed by an evaluation clock pulse in accordance with the operation sequence of the gate explained in Section 2. Bistable buffers clocked on $Phase1$ are required to synchronize inputs to the sum stage of the circuit which is evaluated on $Phase2$ of the clock. The QMOS adder is pipelined at the gate level without necessity for external pipeline latches. This has advantages over conventional pipelining schemes that are commonly used to speed up computation by dividing a combinational block into several sequential stages such that each stage performs a different operation during a particular clock cycle. Conventional pipelining schemes are limited by the added delay and area of pipeline latches. A combinational block consisting of n stages, each having a delay of t_c, will have a total delay of nt_c. If we pipeline this circuit by introducing a latch, having delay t_l, at the output of each stage, the maximum delay of the circuit becomes $(t_c + t_l)$. The throughput of the circuit increases from $1/(nt_c)$ to $1/(t_c + t_l)$ but the latency increases from nt_c to $n(t_c + t_l)$. Also, if a_c is the area of the combinational block; after pipelining, the area of the circuit increases to $a_c + nka_l$, where a_l is the area of a latch and k is the number latches at each stage. If the latch delay t_l is much larger than the stage delay t_c, it places an upper bound on the maximum achievable throughput of the pipelined circuit. Thus, in conventional pipelining schemes, there exist direct tradeoffs between the the area of the pipeline latches and the achievable throughput. However, the use of RTDs and MOSFETs in designing self-latching circuits improves the performance of these gate-level pipelined circuits conventional pipelined circuits because the latch delay, $t_l = 0$. Also, if latency is not of concern, as is the case in signal processing systems and communication systems, each logic gate can operate in the bistable mode resulting in maximum possible throughput.

The QMOS bistable full adder uses 5 RTDs and 20 n-type MOSFETs. To convert a standard 24-MOSFET static CMOS adder to a similar gate-level pipelined adder, we would require an additional 40 transistors for 5 latches required for the carry, the sum and the three input signals. The addition of these latches would provide the pipelining advantage in standard CMOS but would increase the stage delay and area due to the additional latches. Thus, the QMOS logic family has advantage over conventional CMOS logic in terms of greater circuit compactness and improved speed. The application of such gate-level pipelined logic to system design is discussed in the following section.

4 System Applications of QMOS Logic

Communication systems and signal processing systems represent two application areas where deeply pipelined systems are effective because of the volume and similarity of computational requirements at each clock cycle. These systems have minimal data dependence between active computations in pipeline stages and hence are ideal candidates for implementation using gate-level pipelined QMOS logic. We explain a typical system application of QMOS logic using a parallel correlator as an example. Typically, in spread spectrum systems, a parallel correlator computes the correlation of the incoming data stream with a pre-determined PN sequence of a fixed length. This correlation value is used to estimate the output data. For secure and noise-free communication, a long PN sequence length is desirable, resulting in reduced symbol transmission rate. To maintain a reasonable transmission rate, it is therefore necessary to employ high-speed circuits in the receiver. The gate-level pipelined approach of QMOS along with its elimination of pipeline latch overhead facilitates this requirement. The pipelined correlator [7] illustrated in Figure 4 uses a 32-bit latch to hold the PN sequence. The input is a serial bit stream which is fed to a 32-bit shift register. The 32-bit latch and 32-bit shift register are each composed of 64 bistable QMOS inverters. A pair of cascaded bistable inverters each operating on single, separate phases of the two-phase clock form the basic 1-bit latch. The 32-bit raw correlation vector is generated by performing a bitwise XOR operation on the PN sequence latch output and the most recent 32 bits of the sampled signal available at the shift register output. The raw correlation vector is automatically latched at the output of the XOR network due to the use of self-latching gates. This vector forms the input to the pipelined adder network that determines the difference between the number of 1s and 0s in the raw correlation vector at each cycle. The pipelined adder network consisting of 26 nanopipelined full adders, 11 nanopipelined half adders, and 36 bistable inverters, illustrated in Figure 5, forms the core of the parallel correlator. The adders used in the design have complemented sum and carry outputs in order to reduce pipeline latency. The circuit performs eighteen stages of addition to generate a 7-bit result at each clock cycle that is the difference between the number of 1s and number of 0s in the correlation vector. Since the seven bits of the adder network output are not simultaneously

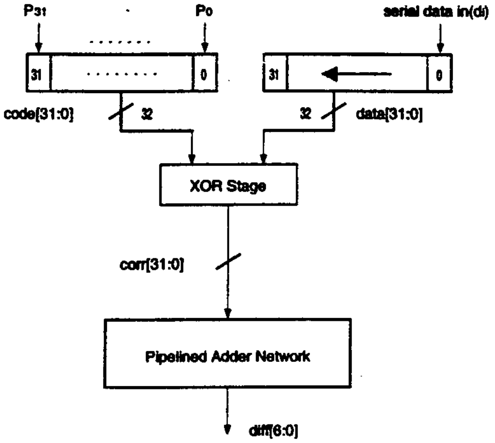


Figure 4: Block diagram of a 32-bit parallel correlator

generated, bistable inverters are required to synchronize the bits such that all seven bits of a correlation appear in order at the output of the correlator. System applications of QMOS logic have been explored for circuits such as pipelined carry-save multipliers, phase accumulators for direct digital frequency synthesizers and turbo-code decoders in digital communication systems. The advantages of QMOS stem from compactness of the logic, inherent bistability and high switching speed of the RTDs, and elimination of bulky and slow PMOS devices. Table 1 compares QMOS with dynamic and static CMOS in terms a normalized area-power-delay metric for various circuits.

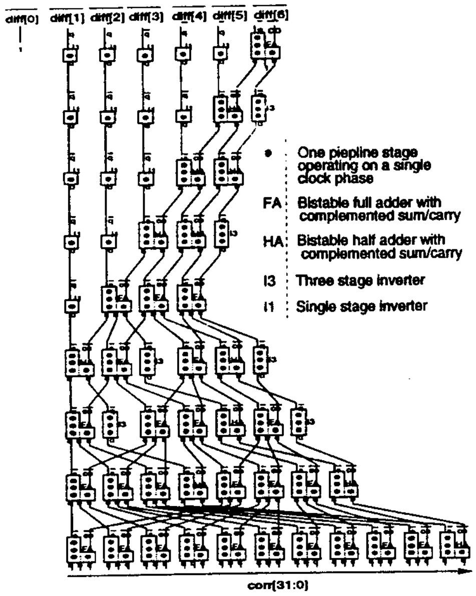

$\overline{\text{diff[0]}}$
$\overline{\text{diff[1]}}$
$\overline{\text{diff[2]}}$
$\overline{\text{diff[3]}}$
$\overline{\text{diff[4]}}$
$\overline{\text{diff[5]}}$
$\overline{\text{diff[6]}}$
• : One piepline stage operating on a single clock phase
FA : Bistable full adder with complemented sum/carry
HA : Bistable half adder with complemented sum/carry
I3 : Three stage inverter
I1 : Single stage inverter
corr[31:0]

Figure 5: 32-bit pipelined adder network

Table 1: Comparison of QMOS with CMOS

Circuit	Static CMOS	Dynamic CMOS	QMOS
Bistable Inverter	14.2	3.9	1
Pipelined Adder	24.9	5.7	1
32-bit Correlator	33.1	20.2	1

5 Conclusions

A design methodology for a new logic family using RTDs and MOSFETs has been presented. The proposed logic family has advantage over conventional CMOS logic in terms of circuit compactness and improved power-delay product. Basic bistable logic gates have been designed in the proposed logic family and their operation has been verified through simulation. The design of complex functions such as a pipelined full adder has been demonstrated. The bistable nature of the circuits eliminates pipeline latch overhead of area, power, and delay, leading to higher throughput in pipelined systems as demonstrated in the design of a 32-bit parallel correlator. In light of this, the proposed logic family has possible applications in high speed communication systems and signal processing systems where a large fraction of the devices are switching at a given time and pipeline latency is not a major concern.

References

[1] F. Capasso, S. Sen, F. Beltram, L. M. Lunardi, A. S. Vengurlekar, P. R. Smith, N. J. Shah, R. J. Malik, and A. Y. Cho, "Quantum functional devices: resonant-tunneling transistors, circuits with reduced complexity and multiple-valued logic," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2065–2082, October 1989.

[2] P. Mazumder, S. Kulkarni, M. Bhattacharya, and A. González, "Circuit design using resonant tunneling diodes," in *Proceedings of the Eleventh International Conference on VLSI Design*, January 1998.

[3] A. C. Seabaugh, A. H. Taddiken, E. A. Beam III, J. N. Randall, Y. C. Kao, and B. Newell, "Co-integrated resonant tunneling diode and heterojunction bipolar transistor full adder," in *International Electron Devices Meeting*, pp. 419–422, 1993.

[4] H. Chan, S. Mohan, P. Mazumder, and G. Haddad, "Compact multiple-valued multiplexers using negative differential resistance devices," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 1151–1156, August 1996.

[5] K. Maezawa and T. Mizutani. "A new resonant tunneling logic gate employing monostable-bistable transition," *Japanese Journal of Applied Physics*, vol. 32, pp. L42–L44, January 1993.

[6] M. Bhattacharya and P. Mazumder, "Spice simulation of circuits containing resonant tunneling diodes," in *Proceedings of the European Conference on Circuit Theory and Design*, 1999.

[7] S. Kulkarni, P. Mazumder, and G. I. Haddad, "A 32-bit ultrafast parallel correlator using resonant tunneling devices," in *International Conference on Integrated Micro/Nanotechnology for Space Applications*, 1995.

Appendix 48

PAPER *Special Issue on Integrated Electronics and New System Paradigms*

Design and Analysis of Resonant-Tunneling-Diode (RTD) Based High Performance Memory System

Tetsuya UEMURA[†a)] *and* **Pinaki MAZUMDER**[††], *Nonmembers*

SUMMARY A resonant-tunneling-diode (RTD) based sense amplifier circuit design has been proposed for the first time to envision a very high-speed and low-power memory system that also includes refresh-free, compact RTD-based memory cells. By combining RTDs with n-type transistors of conventional complementary metal oxide semiconductor (CMOS) devices, a new quantum MOS (Q-MOS) family of logic circuits, having very low power-delay product and good noise immunity, has recently been developed. This paper introduces the design and analysis of a new QMOS sense amplifier circuit, consisting of a pair of RTDs as pull-up loads in conjunction with n-type pull-down transistors. The proposed QMOS sensing circuit exhibits nearly 20% faster sensing time in comparison to the conventional design of a CMOS sense amplifier. The stability analysis done using phase-plot diagram reveals that the pair of back-to-back connected static QMOS inverters, which forms the core of the sense amplifier, has meta-stable and unstable states which are closely related to the I-V characteristics of the RTDs. The paper also analyzes in details the refresh-free memory cell design, known as tunneling static random access memory (TSRAM). The innovative cell design adds a stack of two RTDs to the conventional one-transistor dynamic RAM (DRAM) cell and thereby the cell can indefinitely hold its charge level without any further periodic refreshing. The analysis indicates that the TSRAM cell can achieve about two orders of magnitude lower stand-by power than a conventional DRAM cell. The paper demonstrates that RTD-based circuits hold high promises and are likely to be the key candidates for the future high-density, high-performance and low-power memory systems.

key words: *resonant-tunneling diode, DRAM, sense amplifier, instability, refresh-free*

1. Introduction

The present large demand for high-speed and low-power memory systems has mainly fueled by the explosive growth of large-scale computing and communication systems. Over the past few decades, device dimensions have progressively diminished using the well-known device scaling law in order to meet the increasing needs for higher computing and communication speeds as well as lower energy consumption of pervasive communication and portable electronic systems. Memory devices, especially dynamic random access memories (DRAMs), owing to the tiny size of its memory cell consisting of a vertically integrated capacitor and a pass transistor, define the cutting edge of the VLSI technology. Taking advantages of device scaling, the DRAM industry has relentlessly quadrupled the memory chip size every three years or so, since their first introduction in the early 1970's. The DRAM chip integration density has recently reached the gigabit mark and is projected to grow to 64 Gbits before the device scaling reaches its final limits somewhere near 2010 [1].

This growth potential of DRAM chips, however, is possible if the industry can grapple with the concomitant problems of rising chip power consumption and diminishing improvement in chip access time. As the density of integration increases, the accompanying increase in *RC* delays of longer bit and word lines reduce the memory read/write access times. Also, larger memory chips with billion transistors require significantly more energy consumption for cell refreshing. Furthermore, increased junction leakage currents associated with deeper sub-micron CMOS technology not only increases the refresh power dissipation of a gigabit DRAM chip, but it also mandates the use of more reliable cells with higher storage capacitance.

Since these are intrinsic problems of CMOS technology emanating from progressive device scaling, alternative avenues must be explored to incorporate radical changes in the existing CMOS technology as well as to design new circuit configurations that may obviate cell refreshing and compensate for increasing interconnect delays. Quantum effect devices are envisioned as a viable solution to these two formidable problems that stand as roadblocks for the future growth of DRAM chips. These devices have extremely fast switching speed and fold-back I-V characteristics owing to electron tunneling through quantum barriers. These two features can be utilized to design ultra-fast and complex digital gates which use less active devices and can perform pipelining at the gate level due to the bistability of quantum devices. Though a large number of quantum effect devices have been proposed by the device researchers, resonant tunneling diodes (RTDs) have gained much attention because of their compatibility with many conventional technologies such as HBTs, HEMTs, HFETs and MOSFETs. Such RTD-based emerging technologies hold tremendous potential for dramatic improvements in VLSI circuits [2].

Recently, Wagt, et al. proposed a revolutionizing

Manuscript received January 20, 1999.
Manuscript revised March 16, 1999.

[†] The author is with the Fundamental Research Laboratories, NEC Corporation, Tsukuba-shi, 305-8501 Japan.

[††] The author is with the Department of Electrical Engineering and Computer Science, The University of Michigan, MI. USA.

a) E-mail: uemura@frl.cl.nec.co.jp

Appendix 48

DRAM cell design, called tunneling SRAM (TSRAM), consisting of a pair of RTDs to store the cell information. RTDs, connected in tandem, can hold the charge level of the cell indefinitely replenishing continually the charge loss caused by the cell leakage currents, like the weak inversion current, field current and dark current. Unlike in DRAM cell design, where capacitors are dynamically and periodically refreshed through highly capacitive bit lines, TSRAM cells do not require separate refreshing, and thereby the power consumption of a TSRAM memory chip is considerably lower than a DRAM chip employing external or auto-refreshing [3]. In this paper, a novel RTD-based sense amplifier circuit, called a quantum MOS (QMOS) sense amplifier, is also proposed and its operation is analyzed and verified through SPICE simulation. The sense amplifier is shown to have better response time than a pure CMOS sense amplifier. Finally, the paper demonstrates how by combining RTDs with CMOS devices, low-power and high-speed memory system can be built.

2. RTD-Based Memory System

Figure 1 represents the proposed RTD-based memory system, consisting of QMOS sense amplifier and TSRAM cell. In the TSRAM, a pair of negative differential resistance (NDR) devices ($N3$ and $N4$), such as RTDs or Esaki tunneling diodes (TDs), with their current densities comparable to the cell leakage current, is added between access transistor ($M7$) and storage capacitance (C_S) [3]. The QMOS sense amplifier is formed by the cross-coupled inverter latch. It consists of the RTD pull-up loads ($N1$ and $N2$), NMOS transistors ($M1$ and $M2$), and switching transistors ($M7$, $M8$). The bit-line pair is connected to the drain of $M1$ and $M2$.

Stand-by Mode

In stand-by mode, each cell is isolated from the [illegible] As shown in Fig. 2, the cell voltage stored in the capacitor is latched at either V_H or V_L, which corresponds to logic '*high*' and '*low*,' respectively. In this diagram, i_A and i_B indicate the currents through $N3$ and $N4$, respectively, and i_l is the cell leakage current. Because of this bistable property, the cell can [illegible] hold its information and consequently the [illegible] does not require any periodic refreshing [illegible] in conventional DRAM chips. In order to [illegible] stand-by power consumption, the currents [illegible] $N3$ and $N4$ should be held as low as possible [illegible] current drive capabilities of $N1$ and $N2$, on [illegible] should be as large as those of the $M1$ [illegible] transistors, because the sense amplifier circuit [illegible] the bit-line pair quickly.

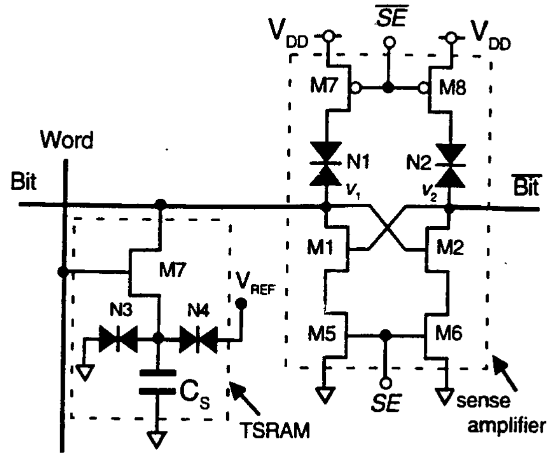


Fig. 1 Proposed RTD-based memory system.

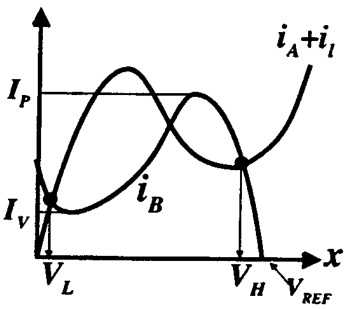


Fig. 2 Load line diagram of TSRAM cell.

2.2 READ/WRITE Operation

In the *READ* operation, the bit lines are precharged to half-V_{DD}. When the access transistor ($M7$), turns *ON*, charge sharing between the bit-line and the accessed cell induces a small voltage difference between the BIT and $\overline{BIT}$ nodes. Once the voltage difference is established, control signals (SE and $\overline{SE}$), activate the cross-coupled latch, and the small voltage difference is amplified to push the BIT voltage to either V_{DD} or 0, depending on the sign of initial voltage difference. Although the stored data is destroyed due to the charge sharing, it is restored by the sense amplifier like in a normal DRAM read operation. The RTD-pair circuit in the TSRAM has no influence on this *READ* operation, because the current drive capability of $N3$ and $N4$ is much smaller than the drive capabilities of the access transistor and the sense amplifier circuit.

In the *WRITE* operation, the input data is written to the cell through the access transistor by charging the bit line connected to the selected cell to either $V_H + V_T$ or $V_L - V_T$, depending on the input data, where V_T is the threshold voltage of the access transistor. The RTD pair circuit, $N3$ and $N4$, has no influence on the *WRITE* operation.

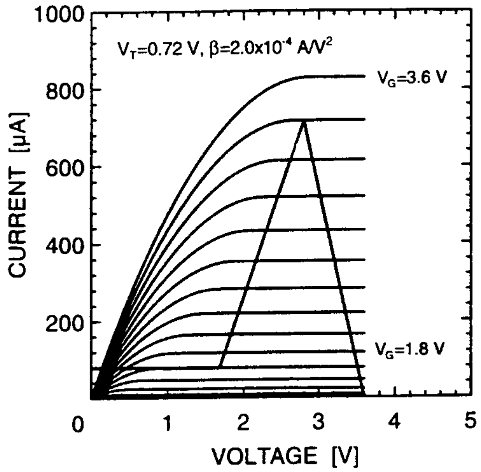


Fig. 3 Load line diagram for QMOS inverter circuit.

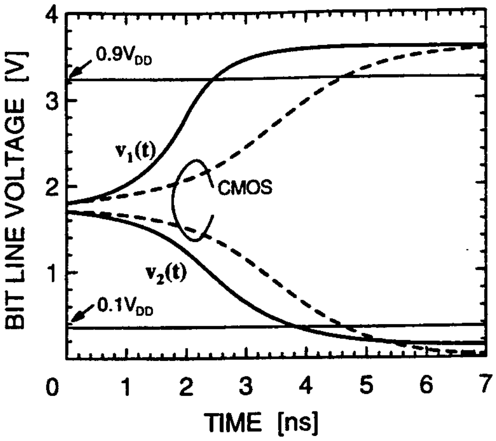


Fig. 4 Simulation results for QMOS and CMOS sense amplifier circuit.

3. Analysis of RTD-Based Memory System

In this section, the circuit performance of the RTD-based memory system is analyzed in detail and compared with the conventional DRAM system.

3.1 Operation Speed

As described in the previous section, the RTD pair circuit has no influence on the READ/WRITE operation. Hence, the access time is determined mainly by the time needed for charging/discharging of the bit line through its sense amplifier. Figures 3 and 4 show a load line diagram of a QMOS inverter used for this analysis, and the SPICE simulation result of sensing time of the coupled QMOS inverter, respectively. The RTDs with $I_P = 718\,\mu$A, $V_P = 0.8$ V, $I_V = 77.4\,\mu$A, $V_V = 1.92$ V, and $V_{SEC} = \infty$, and the NMOS transistors with $\beta = 2.0 \times 10^{-4}$ A/V^2 and $V_T = 0.72$, are used, where the I_P, V_P, I_V, V_V, and V_{SEC} represent a peak current, a peak voltage, a valley current, a valley voltage and a second peak voltage for the RTD characteristics, respectively. The β and V_T denote the gain and threshold voltage, respectively, for the NMOS transistor. The initial value of the bit line voltage is 1.8 V and 1.7 V, and its capacitance is assumed to be 0.5 pF.

The dashed line in Fig. 4 indicates the sensing time of CMOS sense amplifier shown for the comparison. The NMOS characteristics are assumed to be the same in the both sense amplifier circuits. The PMOS characteristics of the CMOS sense amplifier is assumed to be symmetric to the NMOS characteristics ($\beta_P = \beta_N$, $V_{TP} = -V_{TN}$). The sensing time, defined as a time at which both conditions $v_1 > 0.9V_{DD}$ and $v_2 < 0.1V_{DD}$ are simultaneously satisfied, is about 20% faster than that of the CMOS. Since the RTD, as a pull-up load device, can drive more current than an equivalent PMOS load device, the sensing time improves in the Q-MOS sense amplifier. As the current drive capability of the pull-up load is larger, the charging time of the bit-line capacitance becomes shorter. This fast charging time induces the reduction of discharging time of other bit-line capacitance also, because the NMOS transistor of the discharging circuit quickly turns ON due to the cross-coupled connection.

Another improvement of speed performance in this system comes from the refresh-free operation due to the TSRAM configuration. Unlike in memory chips that require periodic refreshing, in TSRAM the *READ/WRITE* operation can be done at any time.

3.2 Stand-by Power Consumption

Since the switching transistors, $M5 - M8$, inactivate the cross-coupled latch during the stand-by mode, the sense amplifier circuit consumes no static power. The static power of the TSRAM cell, on the other hand, is determined by the cell leakage current and latching current. As shown in Fig. 2, in order to guarantee the bistability, the peak currents of $N3$ and $N4$ should be satisfied by the condition given by:

$$I_P > I_V + I_l, \tag{1}$$

where I_P and I_V are the peak and valley currents of $N3$ and $N4$, and I_l is the cell leakage current. Here the latching current of the RTDs is assumed to be I_V. Because of their extremely small sizes, DRAM cells are very sensitive to chip defects. As the density of DRAM increases, the chip yield could be very low. Let γ be the ratio between the maximum cell leakage current (I_l^{max}) and the average leakage current (I_l). The value of γ as high as 50 were reported in the 16 Mb DRAM, although the fraction of the *bad* cell is as small as 10^{-6} [4]. Since Eq. (1) should be satisfied for all cells, the valley current should be larger than the I_V^{min}, which is given by,

$$I_V^{min} = \frac{\gamma}{PVCR - 1} I_l. \tag{2}$$

where $PVCR$ is the peak-to-valley-current-ratio of the RTDs. The stand-by power per cell is given by,

$$\begin{aligned} P_{TSRAM} &= (I_V + I_l)V_{DD} \\ &= \left(\frac{\gamma}{PVCR-1}+1\right) I_l V_{DD} \end{aligned} \qquad (3)$$

3.2.1 Comparison with DRAM

The most significant difference between the TSRAM and DRAM is the stand-by operation. The DRAM needs a repeated refresh operation to hold the cell information, while the TSRAM has no refresh operation due to the data latching by RTD circuit. In this section, power consumption of the DRAM refresh operation and that of the TSRAM stand-by operation are compared.

As shown in the Appendix, the power consumption of the DRAM due to the refresh operation is given by,

$$P_{DRAM} = \frac{1}{t_R}(C_B + C_S)V_{DD}\left(\frac{1}{2}V_{DD}, -\Delta v\right) \qquad (4)$$

where t_R is the cell retention time, Δv is the bit-line voltage change due to charge sharing. When the detection limit of the sense amplifier circuit, V_r, is ideally 0, the above equation simply becomes

$$P_{DRAM} = \gamma\left(1 + \frac{C_B}{C_S}\right) I_l V_{DD} \qquad (5)$$

As γ increases, the power for both DRAM and TSRAM increases almost linearly. The ratio between P_{DRAM} and P_{TSRAM} is proportional to the C_B/C_S as well as the $PVCR$. Because of the factor, C_B/C_S, which is typically more than 10, the P_{DRAM} becomes much larger than the P_{TSRAM}. In other words, one of the reasons for larger power consumption in DRAM emerges from the fact that it necessitates charging the bit line having much larger capacitance than a cell for each refresh operation. The increase in the $PVCR$ value also increases the power ratio, P_{DRAM}/P_{TSRAM}.

Furthermore, when V_r has a finite value, the power, P_{DRAM} increases, resulting in larger power ratio. Figure 5 shows the comparison between power consumptions in TSRAM (Eq. (3)) and in DRAM (Eq. (4)) as a function of V_r, when $V_{DD} = 3.6$ V, $C_S = 27$ fF, $C_B = 270$ fF, $I_l = 1$ fA, and $\gamma = 50$. The P_{DRAM} is normalized by the $P_{TSRAM} = 0.048$ pW. The $PVCR$ of TSRAM is assumed to be 5. Despite the value of P_{DRAM} is slightly underestimated in the calculation, it is still about two orders of magnitude higher than the power consumption of a TSRAM, in the practical V_r range ($\simeq 0.1$ V).

It should be noted that considering the effect of variation in the RTD currents, the power consumption of the TSRAM is increased. This variation can be treated as an increase of the γ value in Eq. (3). The power consumption increases almost linearly against the γ. However, the redundancy cell scheme, which is widely used in the recent high density DRAM, can avoid a significant increase of γ value.

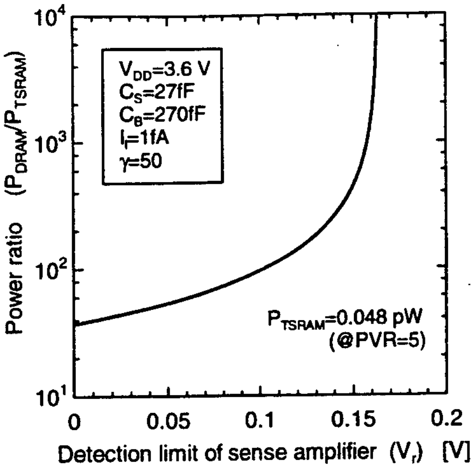


Fig. 5 Comparison of power consumption for the DRAM and TSRAM.

3.3 Stability

The sense amplifier detects the small bit-line voltage difference and accordingly pushes the bit-line voltage to V_{DD} or 0 V, as shown in Fig. 4. The sensing time as well as the convergence value are sensitive to the initial conditions. In this section, the stability of the QMOS sense amplifier against the fluctuation of initial bit-line voltage is discussed by using phase diagram.

The circuit equation of Fig. 1 during the sensing is given by

$$C_B\frac{dv_1}{dt} = I_{RTD}(V_{DD} - v_1) - I_{TR}(v_2, v_1) \qquad (6)$$

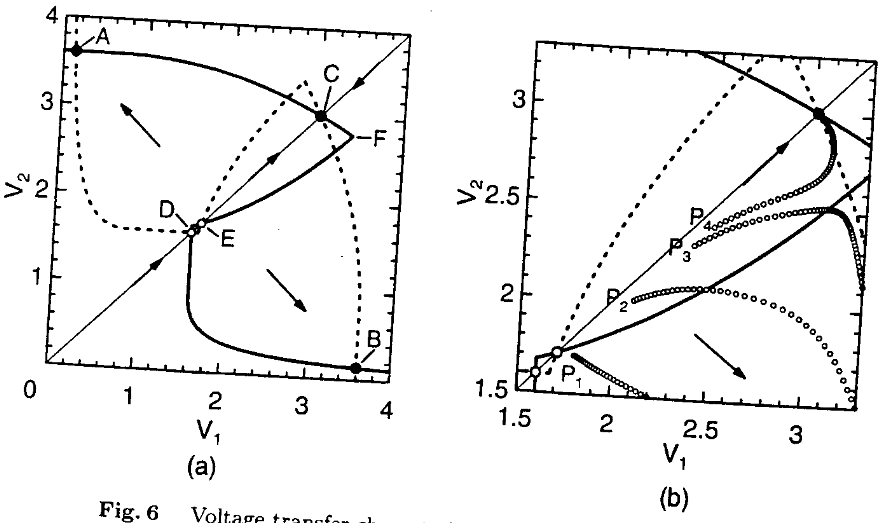
$$C_B\frac{dv_2}{dt} = I_{RTD}(V_{DD} - v_2) - I_{TR}(v_1, v_2) \qquad (7)$$

where C_B is the bit-line capacitance, v_1 and v_2 are BIT and $\overline{BIT}$ voltages. The $I_{RTD}(v)$ and $I_{TR}(v_g, v_{ds})$ indicate the current-voltage characteristics of the RTDs ($N1$ and $N2$) and the NMOS transistors ($M1$ and $M2$), respectively, which are given by

$$I_{RTD}(v) = \begin{cases} \frac{I_P}{V_P}v & (0 \le v < V_P) \\ I_P - \frac{I_P - I_V}{V_V - V_P}(v - V_P) & \\ & (V_P \le v < V_V) \\ I_V + \frac{I_P - I_V}{V_{SEC} - V_V}(v - V_V) & (V_V \le v) \end{cases} \qquad (8)$$

and

$$I_{TR}(v_g, v_{ds}) = \begin{cases} 0 & (v_g < V_T) \\ \beta[(v_g - V_T)v_{ds} - v_{ds}^2/2] & \\ & (v_{ds} < v_g - V_T) \\ \frac{\beta}{2}(v_g - V_T)^2 & (v_{ds} \ge v_g - V_T) \end{cases} \qquad (9)$$



(a)

(b)

Fig. 6 Voltage transfer characteristics of QMOS inverter pair.

where the I_P, V_P, I_V, V_V, and V_{SEC} represent the peak current, the peak voltage, the valley current, the valley voltage and the second peak voltage for the RTD characteristics, respectively. The β and V_T are gain and threshold voltages, respectively, for the NMOS transistor.

Figure 6 shows voltage transfer characteristics (VTC) of each inverter. The solid and dashed VTC lines indicate the solution for $dv_2/dt = 0$ and $dv_1/dt = 0$, respectively. The intersection points, therefore, represent the equilibrium points of Eqs. (6) and (7). Unlike the CMOS sense amplifier. the QMOS sense amplifier circuit has three stable points (A, B, and C), one metastable point (D), and one unstable point (E), due to its Z-shaped VTC. as shown in Fig. 6.

The solution of Eqs. (6) and (7) for a given initial condition is discussed using phase diagram. The dotted lines in Fig. 6(b) indicate the trajectory curves of Eqs. (6) and (7) for different initial conditions. From the point $P_1(1.8, 1.7)$, where $dv_1/dt > 0$ and $dv_2/dt < 0$, the solution directly goes to the stable point B (Fig. 4). From $P_2(2.0, 1.9)$ or $P_3(2.4, 2.3)$, where both dv_1/dt, and dv_2/dt, > 0, the trajectory once goes upward, and when it reaches to the solid VTC curve, it turns to the downward, resulting in going to the point B. When it starts from the point $P_4(2.5, 2.4)$, however, the solution reaches to the dashed VTC curve, resulting in convergence to the point C, exhibiting an erroneous operation of the sense amplifier.

This erroneous behavior can be controlled by optimizing the RTD parameters. If the point F on the solid VTC curve, corresponding to the peak position of RTD characteristics, is moved closer to the line of $v_2 = v_1$, the instability is reduced. The v_1 and v_2 values at point F, which is determined by the peak current and voltage of the RTD characteristics, are given by

$$v_1|_F = V_T + \sqrt{\frac{2I_P}{\beta}}$$

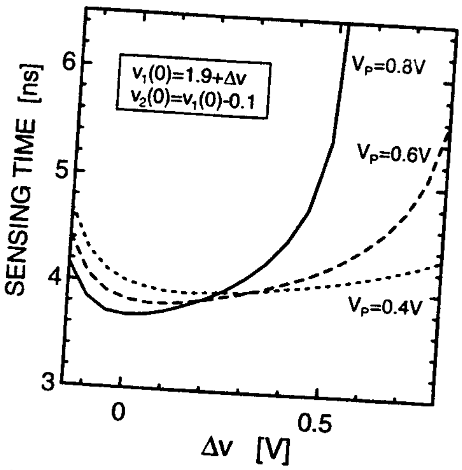


Fig. 7 Initial voltage fluctuation dependence of the sensing time.

$$v_2|_F = V_{DD} - V_P \tag{10}$$

From the above equations, when I_P is decreased, the point F is moved to left. When V_P is decreased, on the other hand, the point F is moved upward. Therefore, the decrease of I_P and/or V_P can reduce the instability. Figure 7 shows an initial condition dependence of the sensing time for three different peak voltage values (V_P =0.8, 0.6, and 0.4 V). The initial condition for $v_1(t)$ and $v_2(t)$ are assumed to be,

$$\begin{aligned} v_1(0) &= v_2(0) + 0.1 \\ v_2(0) &= V_{DD}/2 + \Delta v \end{aligned} \tag{11}$$

where Δv indicates the voltage fluctuation during the bit-line precharge cycle. As shown in Fig. 6 (b), when Δv equals to 0.6 V at $V_P = 0.8$ V, which corresponds to the point P_4 in Fig. 6 (b), the sensing time becomes infinity due to the erroneous operation. Figure 7 indicates that the reduction of V_P can increase the normal operation margin against the initial voltage fluctuation without degrading the sensing speed.

3.4 Memory Cell Capacitor

One of the most critical challenges which gigabit density DRAM designers confront now is the memory cell capacitance. Memory cell capacitance is a critical parameter that determines the sensing signal voltage, sensing speed, data retention time, and vulnerability for soft error. From Eq. (A·3) and Eq. (A·4) in the Appendix, the minimum capacitance value, C_S^{min}, is given by,

$$C_S^{min} = \frac{V_r C_B + I_l t_R}{V_{DD}/2 - V_r} \tag{12}$$

Because of the lower supply voltage and increased junction leakage current due to high doping density, the higher memory cell capacitance will be needed in the gigabit density DRAMs [1].

The TSRAM cell can alleviate the cell capacitor limitation to some extent. The required minimum cell capacitance of the TSRAM cell is independent of the cell leakage current, and is given by

$$C_{S\,TSRAM}^{min} = \frac{V_r C_B}{V_{DD}/2 - V_r} \tag{13}$$

This condition shows that the bit-line voltage shift due to charge sharing during a *READ* operation should be larger than the detection limit of the sense amplifier. In addition, the intrinsic capacitance of $N3$ and $N4$ also works as a node capacitance.

3.5 Cell Area

Disadvantages of the TSRAM include slight increase in cell area and the addition of one extra bias line in the cell layout. Since the current level needed for the RTDs is small, the area could be reduced by adjusting the peak current density. For example, in order to obtain the peak current density of 0.1 pA, which is about 100 times larger than the typical cell leakage current, the RTD with its peak current density of 10^{-3} A/cm^2 occupies only a 0.01 μm^2. The area of the RTD is limited by the size of the contact hole. By forming the RTD pairs on the drain region of the access transistor, the cell area becomes minimum. Since the TSRAM cell needs a contact between upper RTD and power supply (V_{REF}) and a contact between lower RTD and cell plate in addition to the bit-line contact, the word line contact, and the cell capacitance contacts, the theoretical limitation of the cell area is estimated to be $10\lambda^2$, where λ is the minimum technological feature size. An example of the TSRAM cell layout is shown in Fig. 8. The area of TSRAM cell is larger than that of the DRAM whose theoretical lower bound is $6\lambda^2$ (open bit-line scheme). However, recent needs for high cell capacitance in the gigabit density DRAMs tends to increase the cell area from their theoretical bounds. Therefore, the difference for both cell areas is not significantly large.

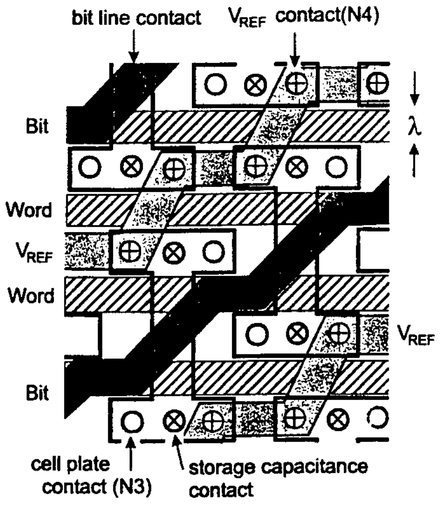


Fig. 8 Cell layout of TSRAM.

3.6 RTD Configuration

In our RTD-based memory system, two kinds of RTDs with different current densities are necessary. The current density of the RTD used in the TSRAM cell should be as low as possible to reduce the stand-by power consumption, while that in the sense amplifier should be as large as possible to increase the sensing speed. The fabrication of two kinds of RTDs at different places, however, induces an area penalty, which in particular is a serious problem in the TSRAM cell design. In this section, we would like to briefly discuss how to integrate two kinds of RTDs without increasing the area. Since the RTD is a vertical transport device, several RTDs can be stacked vertically. In this system, the RTD with smaller current density should be stacked on that with larger current density. In the sense amplifier circuit, where the lower-side RTD is used, the upper-side RTD is removed to make contact directly to the lower-side RTD. In the TSRAM cell, on the other hand, where the upper-side RTD is used for the data latching, the contact is made to the upper-side RTD. Because of much higher current density of the lower-side RTD, this lower-side RTD works as a resistor with low value.

4. Conclusion

This paper demonstrates the design, analysis and simulation of a novel QMOS sense amplifier circuit consisting of RTD pull-up loads and NMOS transistors. Compared to the conventional CMOS sense amplifiers, the QMOS design exhibits about 20% higher sensing speed due to the larger current drive capability of the RTD load. The instability against the initial condition has been analyzed by drawing phase plot diagram. It has been found that by reducing the peak current and/or peak voltage of the RTDs instability problem can be circumvented.

In addition to a novel sense amplifier design, the paper also analyzes the power consumption of a refresh-free TSRAM cell in comparison with the power budget of a DRAM cell that needs periodic refreshing. The stand-by power consumption of the TSRAM was derived theoretically with reference to the DRAM cell dissipation, and it was observed that the TSRAM cell power consumption was about two orders of magnitude lower than DRAM. Further, because of its refresh-free nature, the TSRAM does not suffer from the high cell capacitance requirement. Due to increased junction leakage current in deep sub-micron CMOS technologies, DRAM cells require larger capacitances to optimize the design between cell stability, power consumption, cell area, soft error rate, etc. Since RTDs can be vertically integrated and can hold the charge level indefinitely, TSRAM design methodology may alleviate the critical cell design problems in the future multi-gigabit DRAM chips.

In order to meet the increasing needs for higher speeds and lower power budget in portable computing and communication systems, quantum-MOS (QMOS) technology with low power-delay product and high noise immunity may provide a viable solution. QMOS combines the integration level of CMOS technology with the bistable latching property of a tunneling device and thereby it can be employed for gigabit memory technology. Conventional CMOS technology is confronting major impediments in the design of high-performance and low-power DRAM chips. These roadblocks are due to the intrinsic limitations of the CMOS technology and a viable solution lies in developing a CMOS-based new technology that can radically change the conventional circuit design principle. The paper reveals the promises of QMOS being one such viable technology that can be co-integrated with the existing CMOS technology for enhancing the horizon of VLSI technology to 50-nm scaling.

References

[1] K. Kim, C-G. Hwang, and J.G. Lee, "DRAM technology perspective for Gigabit era," IEEE Trans. Electron Devices, vol.45, pp.598–608, 1998.

[2] P. Mazumder, S. Kulkarni, M. Bhattacharya, J.P. Sun, and G.I. Haddad, "Digital circuit application of resonant tunneling devices," Proc. IEEE, vol.86, pp.664–686, 1998.

[3] P. van der Wagt, A. Seabaugh, and E. Beam III, "RTD/HFET low standby power SRAM gain cell," 1996 IEDM Tech. Dig., pp.425–428, 1996.

[4] H. Yamauchi, T. Suzuki, A. Sawada, T. Iwata, T. Tsuji, M. Agata, T. Taniguchi, Y. Odake, K. Sawada, T. Ohnishi, M. Fukumoto, T. Fujita, and M. Inoue, "A 20 ns battery-operated 16 Mb CMOS DRAM," 1993 ISSCC Tech. Dig., pp.44–45, 1993.

Appendix A: Refresh Operation

In this appendix, the power dissipation due to the refresh operation of DRAM is analyzed. The refresh operation is made by repeatedly performing forced *READ* operations. Each forced *READ* refreshes all cells connected to the same word line. The cell retention time is limited by the cell leakage current of the worst cell. Let t_R be the cell retention time, which is given by

$$t_R = \frac{C_S}{I_l^{max}}(V_{DD} - V_R^{min}) \qquad (A\cdot 1)$$

where V_R^{min} is the minimum cell voltage when the access transistor is *ON*. The bit-line potential change due to the charge redistribution is given by

$$\Delta v_0 = \frac{C_S}{C_S + C_B}\left(V_R^{min} - \frac{1}{2}V_{DD}\right) \qquad (A\cdot 2)$$

Here it is assumed that the bit-line pair is initially precharged to $V_{DD}/2$. This charge sharing requires the power dissipation in the access transistor. From the condition that $\Delta v_0 > V_r$, where V_r is the detection limit of the sense amplifier, the V_R^{min} in the above equation is given by

$$V_R^{min} = \frac{1}{2}V_{DD} + \frac{C_S + C_B}{C_S}V_r \qquad (A\cdot 3)$$

The node voltage and bit-line potential change of the normal cell with its leakage current of I_l, then, become

$$\begin{aligned} V_R &= V_{DD} - \frac{I_l}{C_S}t_R \\ &= V_{DD} - \frac{1}{\gamma}(V_{DD} - V_R^{min}) \end{aligned} \qquad (A\cdot 4)$$

and

$$\Delta v = \frac{C_S}{C_S + C_B}\left(V_R - \frac{1}{2}V_{DD}\right), \qquad (A\cdot 5)$$

respectively. After charge sharing occurs, the sense amplifier is activated, resulting that the bit line and cell node are charged from $1/2V_{DD} + \Delta v$ to V_{DD}, while the $\overline{\text{bit line}}$ is discharged from $1/2V_{DD}$ to 0. During this (dis)charging process, the sense amplifier dissipates some power.

At the end of the previous active cycle, one bit line is at V_{DD} and the other is at 0 V. Turning on the equalization device, which shorts two bit-line halves together, induces the charge sharing between the bit lines, resulting in an initial precharge level at nearly half V_{DD}. The cell voltage is reduced to V_R due to the cell leakage current. Table A· 1 summarizes the voltage value for the cell, bit line, and $\overline{\text{bit line}}$ at each stage.

The total power dissipation comprises of dissipations due to the access transistor during charge sharing

Table A·1 Voltage value for the cell and bit-line pair.

	cell	bit line	$\overline{\text{bit line}}$
initial	V_R	$\frac{1}{2}V_{DD}$	$\frac{1}{2}V_{DD}$
charge sharing	$\frac{1}{2}V_{DD}+\Delta v$	$\frac{1}{2}V_{DD}+\Delta v$	$\frac{1}{2}V_{DD}$
sensing	V_{DD}	V_{DD}	0
precharging	V_R	$\frac{1}{2}V_{DD}$	$\frac{1}{2}V_{DD}$

process, the sense amplifier circuit during the charging/discharging the bit lines, the equalization transistor during the precharging cycle, and the cell itself due to the leakage current. It is equal to the power drawn from the power supply during the bit-line charging. In order to estimate this value, we assume an ideal case, in which all the current from the power supply is used to charge the bit-line. Under this condition, the power drawn from the power supply is given by

$$\begin{aligned} P_{DRAM} &= \frac{1}{t_R}\int_0^T V_{DD}i(t)dt \\ &= \frac{1}{t_R}(C_B+C_S)V_{DD}\int_{V_{DD}/2+\Delta v}^{V_{DD}} dv \qquad (A\cdot 6) \\ &= \frac{1}{t_R}(C_B+C_S)V_{DD}\left(\frac{1}{2}V_{DD}-\Delta v\right) \end{aligned}$$

where T is the sensing time, $i(t)$ is the charging current flowing from the power supply. From this equation, the Eq. (4) is obtained. And when the V_r is equals to 0, the Eq. (5) is obtained.

Tetsuya Uemura received his B.E., M.E., and Ph.D. degrees in electronic engineering from Kyoto University, Kyoto, Japan, in 1988, 1990, and 1996, respectively. He joined NEC Corporation, Japan, in 1990. During 1997–1998, he worked at the Department of Electrical Engineering and Computer Science of the University of Michigan, Ann Arbor, as a visiting researcher. His current interests are quantum effect devices and their circuit application. Dr. Uemura is a member of the Japan Society of Applied Physics.

Pinaki Mazumder received the Ph.D. degree from the University of Illinois at Champaign, USA in 1987. Presently, he is with the Department of Electrical Engineering and Computer Science of the University of Michigan, Ann Arbor as a Professor. He worked at various places that include BEL, India and AT&T Bell Labs in USA, and he spent his sabbatical leave (1996–97) as a visiting faculty at Stanford University, University of California at Berkeley, and Nippon Telegraph and Telephone Corporation, Japan. His research interests include VLSI testing, physical design automation, and ultrafast circuit design. He has published over 100 papers on these topics in archival journals and proceedings of the international conferences. He is the co-author of two books — *"Testing and Testable Design of High-Density Random Access Memories,"* Kluwer Academic Publishers, and *"Genetic Algorithms for VLSI Design, Layout and Test Automation,"* Prentice Hall, 1999. He is a Fellow of IEEE and member of Sigma Xi, Phi Kappa Phi, and ACM SIGDA.

IEEE TRANSACTIONS ON COMPUTERS, VOL. 47, NO. 9, SEPTEMBER 1998

Multiple-Valued Signed-Digit Adder Using Negative Differential-Resistance Devices

Alejandro F. González, *Student Member, IEEE*, and Pinaki Mazumder, *Senior Member, IEEE*

Abstract—This paper describes a new signed-digit full adder (SDFA) circuit consisting of resonant-tunneling diodes (RTDs) and metal-oxide semiconductor field effect transistors (MOSFETs). The design is primarily based on a multiple-valued logic literal circuit that utilizes the folded-back I-V (also known as negative differential-resistance, NDR) characteristics of RTDs to compactly implement its gated transfer function. MOS transistors are configured in *current-mode* logic, where addition of two or more digits is achieved by superimposing the signals of individual wires being physically connected at the summing nodes. The proposed SDFA design uses redundant arithmetic representation and, therefore, the circuit can perform addition of two arbitrary size binary numbers in constant time without the need for either carry propagation or carry look-ahead. The SDFA cell design has been verified through simulation by an augmented SPICE simulator that includes new homotopy-based convergence routines to tackle the nonlinear device characteristics of quantum devices. From the simulation result, the SDFA cell has been found to perform addition operation in 3.5 nanoseconds, which is somewhat superior to other multivalued redundant arithmetic circuits reported in the literature. The SDFA cell requires only 13 MOS transistors and one RTD, as opposed to the state-of-the-art CMOS redundant binary adder requiring 56 transistors, and to the conventional multivalued current-mode adder consisting of 34 MOS transistors. In order to verify the simulation result, a prototype SDFA cell has been fabricated using MOSIS 2-micron CMOS process and GaAs-based RTDs were connected externally to the MOSFET circuit.

Index Terms—Signed-digit arithmetic, multiple-valued logic, quantum electronic resonant-tunneling circuits.

◆

1 INTRODUCTION

IN a conventional ripple-carry adder, the input carry bit may propagate through the bank of full adders, spanning from the least significant digit to the most significant digit. Therefore, the worst-case propagation delay in a ripple-carry adder is proportional to *n*, the size of the adder. Other approaches, such as carry look-ahead, can reduce the propagation delay to log *n* time at the expense of additional circuits that not only introduce irregularity in chip layout but also render the adder circuit less amenable to comprehensive testing. The signed-digit number system, originally proposed in 1961 by Avizienis [1], can be used in adder circuits to restrict carry propagations only to adjoining cells by eliminating the dependency of the carry output function on the carry input signal. To restrict carry propagation, signed-digit number systems employ redundant representation, in which a number different from zero can be expressed in more than one unique way. In signed-digit adders, it is possible to perform addition of two arbitrary size numbers in constant time, and redundant algorithms can, therefore, help to significantly improve the performance of arithmetic circuits in applications with large operand sizes. Signed-digit systems have been adopted by many researchers and designers in the development of high-performance arithmetic circuits [2], [3], [4], [5], but compact and efficient implementation of the signed-digit adders still remains somewhat elusive with the conventional device technologies.

As MOS technology is rapidly advancing to its physical limits of feature size shrinking, it is of paramount importance that device engineers discover alternative enabling technologies that may employ radically different device transportation phenomena, such as quantum tunneling through multiple barrier structures, single-electron-controlled charge transfer over the coulombian blockade, charge transfer over complex molecular structures, and DNA computing. Among a host of nascent technologies that seem to be extremely promising, quantum electronic resonant-tunneling devices, such as resonant-tunneling diodes (RTDs) [6], resonant-tunneling hot electron transistors (RHETs) [7], resonant-tunneling bipolar transistors (RTBTs) [8], bound-state resonant-tunneling transistors (BSRTTs) [9], etc., are the most mature and appear to be imminently viable for commercial introduction. These devices can operate at room temperature and they are compatible with conventional technologies such as heterojunction bipolar transistors (HBTs) [10], [11], and high-electron mobility transistors (HEMTs) [12], [13].

The nonlinear tunneling characteristics of these devices can be efficiently harnessed to design multiple-valued circuits requiring fewer active devices and less amount of connecting wires between them. Interconnects will predominantly govern the circuit speed in future gargantuan multibillion transistor monolithic integrated chips. Multiple-valued logic (MVL) can alleviate the interconnect delay and routing complexities since multivalued signals convey more information than binary signals, thus requiring less amount of interconnects to transmit similar bandwidth of information [14], [15]. In a signed-digit arithmetic system, multivalued signal levels are used to perform the arithmetic

---

- *The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109. E-mail: mazum@eecs.umich.edu.*

Manuscript received 9 June 1997; revised 25 Nov. 1997.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number 105233.

0018-9340/98/$10.00 © 1998 IEEE

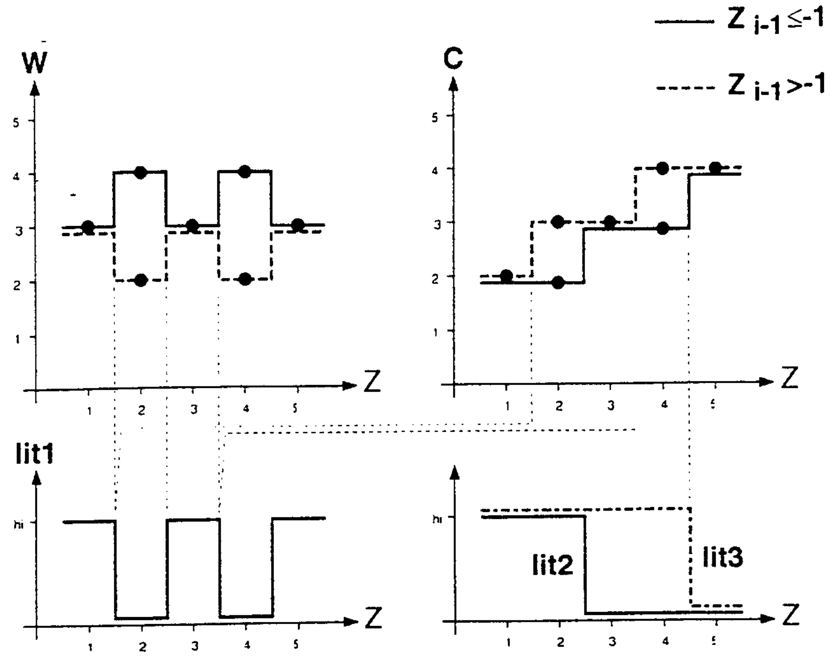


Fig. 2. Transfer function of the SDFA block.

working pair is selected by the value of z_{i-1}. This input signal is used to determine if $c_{i-1} \neq -1$, which indicates when the SDFA cell is allowed to generate an output $w = -1$ without causing invalid s current levels to be produced. If the input z_{i-1} to the previous digit was not considered, then it would be possible to generate $w = -1$ or $w = 1$ when $c_{i-1} = -1$ or $c_{i-1} = 1$, respectively. In these cases, the final sum result would be $s = -2$ or $s = 2$, which are invalid outputs for the selected radix.

Careful observation of the SDFA transfer characteristics shows that three multiple-valued literal signals can be used to describe the adder function. As seen in Fig. 2, literals *lit*1, *lit*2, and *lit*3 contain all the switching information required to define output functions w and c. This observation forms the basis of the proposed SDFA cell design, whose block diagram is depicted in Fig. 3. The input to the system is the wired summation current signal i_z. The three required literal signals are then generated in different blocks of the circuit. The literal signals are used to control switched current sources, which, in turn, synthesize the SDFA transfer functions in the *current output generator* block. Please note that the output block uses input signal $V_{\alpha_{i-1}}$ to determine which of the two sets of transfer characteristics should be used. $V_{\alpha_{i-1}}$ indicates if $z_{i-1} < -1$ and, to our advantage, its behavior is identical to literal *lit*2.

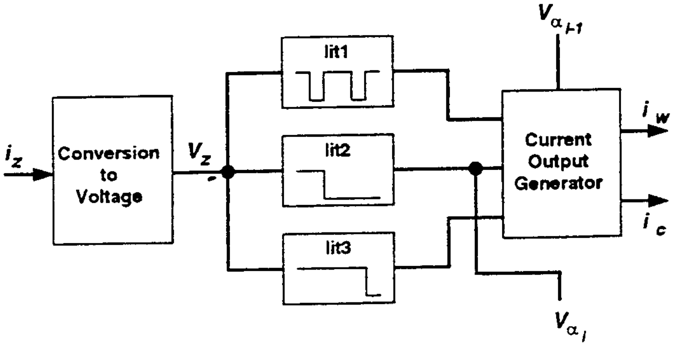
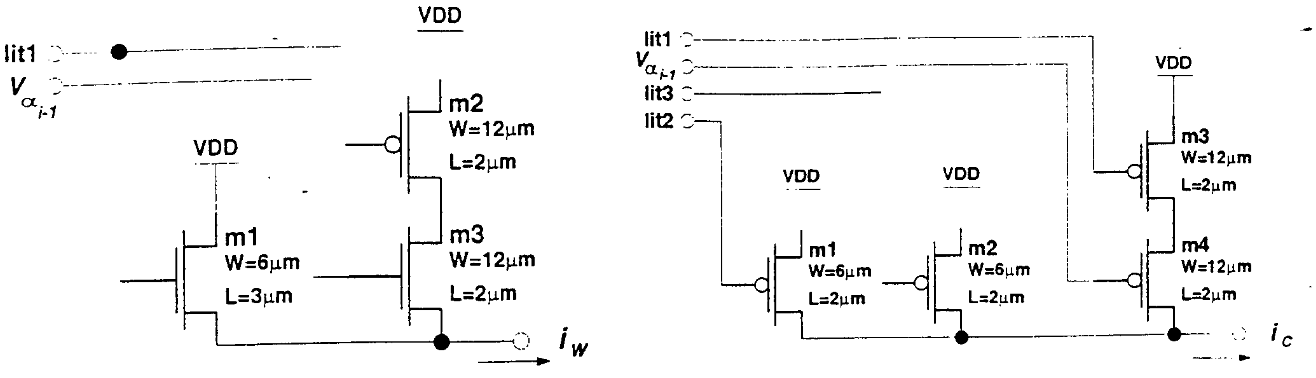


Fig. 3. Block diagram of the proposed SDFA cell.

2.2 Literal Circuit Implementation

Let us begin the description of the SDFA circuit by discussing the implementation of signal *lit*1 using RTDs. Fig. 4 shows the basic circuit used for generating the literal signals. As seen in the figure, the circuit consists of a serial connection of a resistor R_s and an RTD, where the RTD performs as a load element. The input to the circuit is the voltage V_z, which represents the wired summation $z = x + y$. In the case of *lit*1, two RTDs are used in series to obtain an equivalent two-peak characteristics. The CMOS inverter is used to sense the voltage at the node connecting the RTD and the resistor (V_{node}). The proposed literal circuit is similar to the one presented in [14]; however, the proposed design uses a complete CMOS inverter instead of a passive-load inverter.

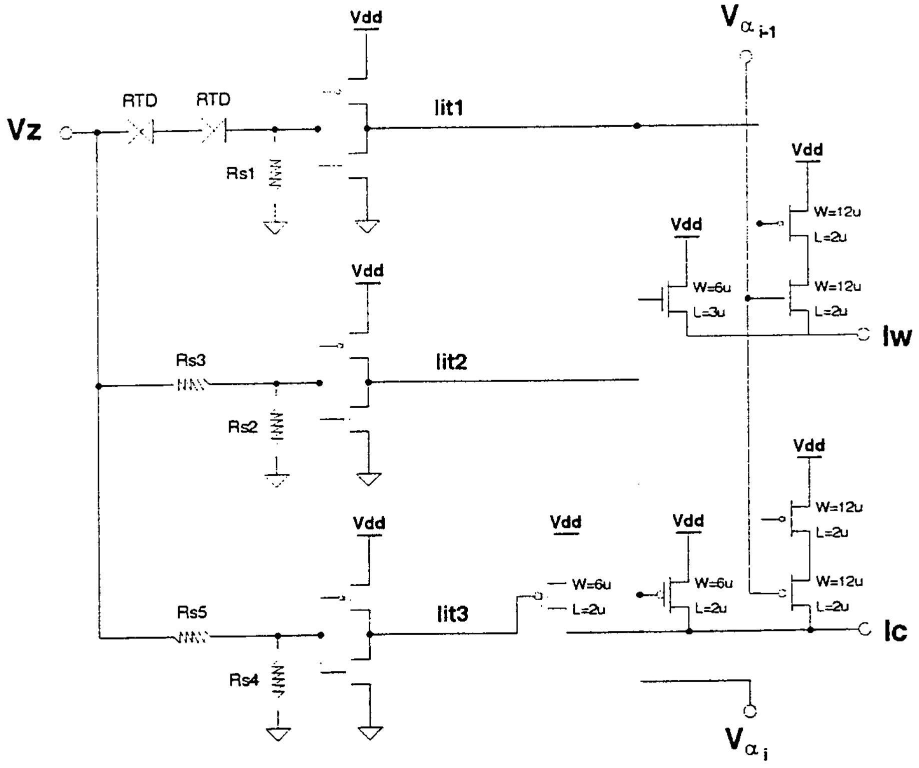
The behavior of the voltage V_{node} is described using the load-line method (depicted in Figs. 5a and 5b). The I-V characteristics of the RTD and the resistor are plotted as the current I that flows through them with respect to the voltage V_{node}. The value of the current I is found at the point of intersection of the I-V curves. The operating point of the circuit changes as the input voltage V_z is increased because the I-V curve of the RTD is moved to the right as V_z increases. There are points where the current I suddenly decreases with increasing V_z due to the folded I-V characteristics of RTDs. Fig. 5b, shows a low-current operating point occurring after a transition from a high-current point (shown in Fig. 5a). Since the voltage V_{node} is proportional to the current I, the transitions in I are directly reflected in V_{node}. Fig. 5c displays the ideal behavior of V_{node} obtained using the load-line method.



VDD
lit1
$V_{\alpha_{i-1}}$
m2
W=12μm
L=2μm
VDD
m1
W=6μm
L=3μm
m3
W=12μm
L=2μm
i_w
lit1
$V_{\alpha_{i-1}}$
lit3
lit2
VDD
m3
W=12μm
L=2μm
VDD
VDD
m1
W=6μm
L=2μm
m2
W=6μm
L=2μm
m4
W=12μm
L=2μm
i_c

(a) (b)

Fig. 6. Circuits for generating the (a) interim sum and the (b) carry output functions.



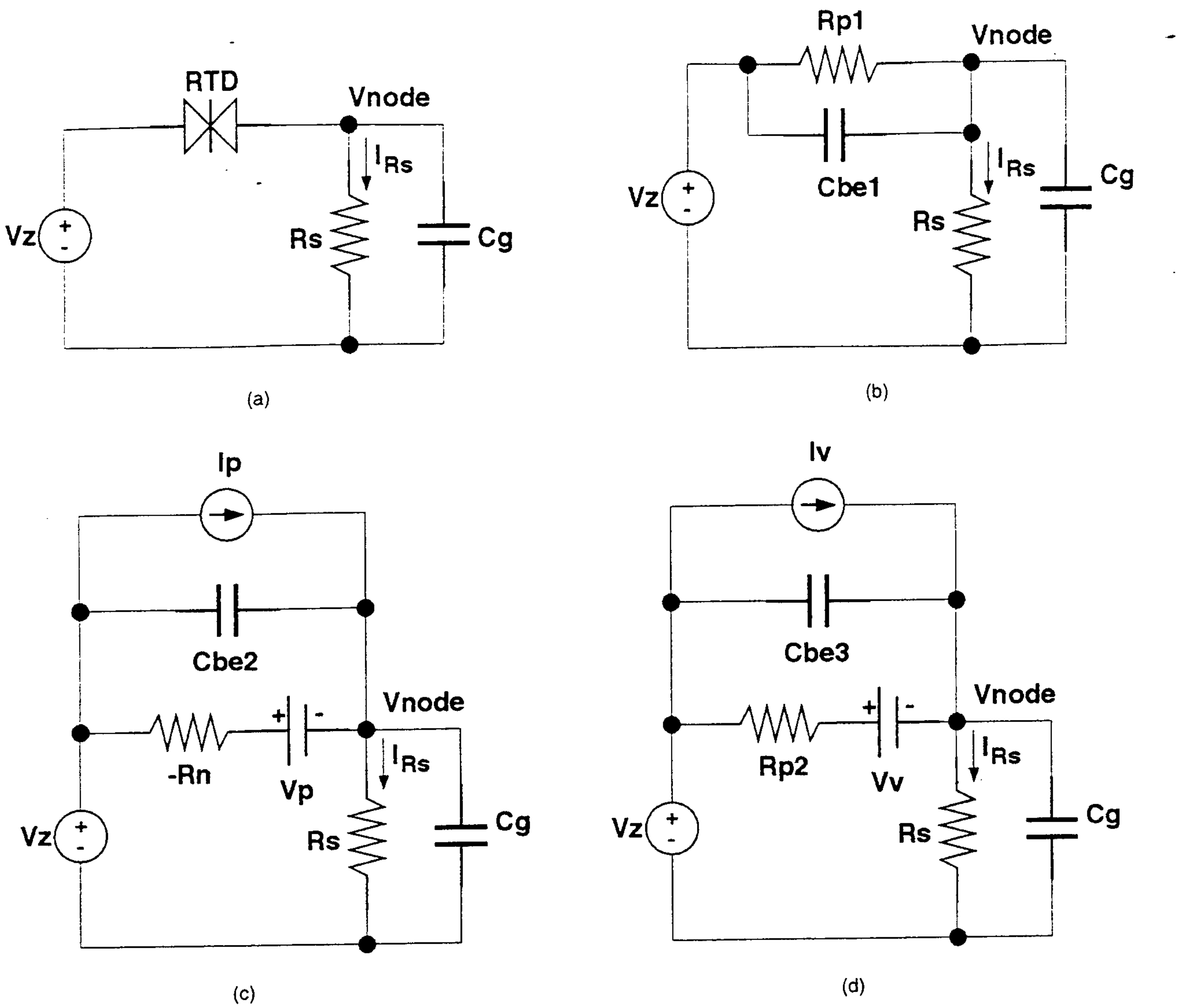
Vdd
$V_{\alpha_{i-1}}$
RTD
RTD
Vz
lit1
Rs1
Vdd
W=12u
L=2u
Vdd
W=6u
L=3u
W=12u
L=2u
Iw
Vdd
Rs3
lit2
Rs2
Vdd
W=12u
L=2u
Vdd
Vdd
Vdd
W=6u
L=2u
W=6u
L=2u
W=12u
L=2u
Rs5
lit3
Ic
Rs4
V_{α_i}

Fig. 7. Complete SDFA cell circuit.

The circuit for implementing function *c* is shown in Fig. 6b. In this circuit, transistors *m*1 and *m*2 produce the basic stairway form of the carry function for the case $z_{i-1} \leq -1$ (as depicted in Fig. 2). Transistors *m*3 and *m*4 are used to generate the correct carry output current, when $z_{i-1} > -1$, by injecting one logic level of current at the operating points where the two cases of the output function *c* are different ($z = 2, 4$). Please note that *lit*1 controls these operating points because they coincide with the zero-level points of this literal signal. Also note that the current output functions generated by the described circuits are not identical to the ideal transfer functions shown in Fig. 2. The difference lies in the current output levels. While the ideal transfer function sweeps current levels 2, 3, and 4, the given output circuits generate signals that sweep levels 0, 1, and 2. This represents no problem at all. Since only positive current signals are used, it is always necessary to perform a shift correction after every wired-addition, and the amount of this correction is smaller if the lowest current levels possible are used. Fig. 7 shows a diagram of the SDFA circuit.

3 TRANSIENT ANALYSIS OF MULTIVALUED LITERAL CIRCUIT

In this section, we analyze the propagation delay of the RTD-based literal circuit. As expected, the propagation delay is a function of the series resistance, R_s, and the RTD characteristics. The analysis yields the best value of the resistance R_s for optimal propagation delay. A piecewise linear RTD model, shown in Fig. 8, is assumed in the analysis.



(a) (b)

(c) (d)

Fig. 9. Circuits used in the resistor-RTD circuit transient analysis. (a) Original circuit. Equivalent circuit in the (b) first PDR region, (c) NDR region, and (d) second PDR region.

second part is done for the NDR region of operation of the RTD, whose equivalent circuit is shown in Fig. 9c. In this part, the delay time $t_{dynamic}$, up to the operating point of interest ($I_{R_s} = (I_p + I_v)/2$), is obtained.

The first part of the analysis (PDR region) is done as follows. Please refer to Fig. 9b for this discussion. The transient response is analyzed by solving the node equations for V_{node} in time. The flow of currents at V_{node} is written as

$$\frac{V_z(t) - V_{node}(t)}{R_{p1}} + C_{be1}\frac{d}{dt}\left(V_z(t) - V_{node(t)}\right) = \frac{V_{node}(t)}{R_s} + C_g\frac{d}{dt}\left(V_{node}(t)\right). \qquad (3)$$

As mentioned previously, the input signal V_z is assumed to be a ramp with slew rate S_R, that is,

$$V_z(t) = S_R t. \qquad (4)$$

Solving the linear differential equation (3) with (4) and the initial condition $V_{node}(0) = 0$ yields the following expression for V_{node}:

$$V_{node}(t) = \frac{R_s S_R t}{R_{p1} + R_s} + \frac{S_R R_{p1} R_s}{R_{p1} + R_s}\left[C_p - \frac{R_s\left(C_g + C_{be1}\right)}{R_{p1} + R_s}\right]\left(1 - e^{-\alpha_1 t}\right), \qquad (5)$$

where

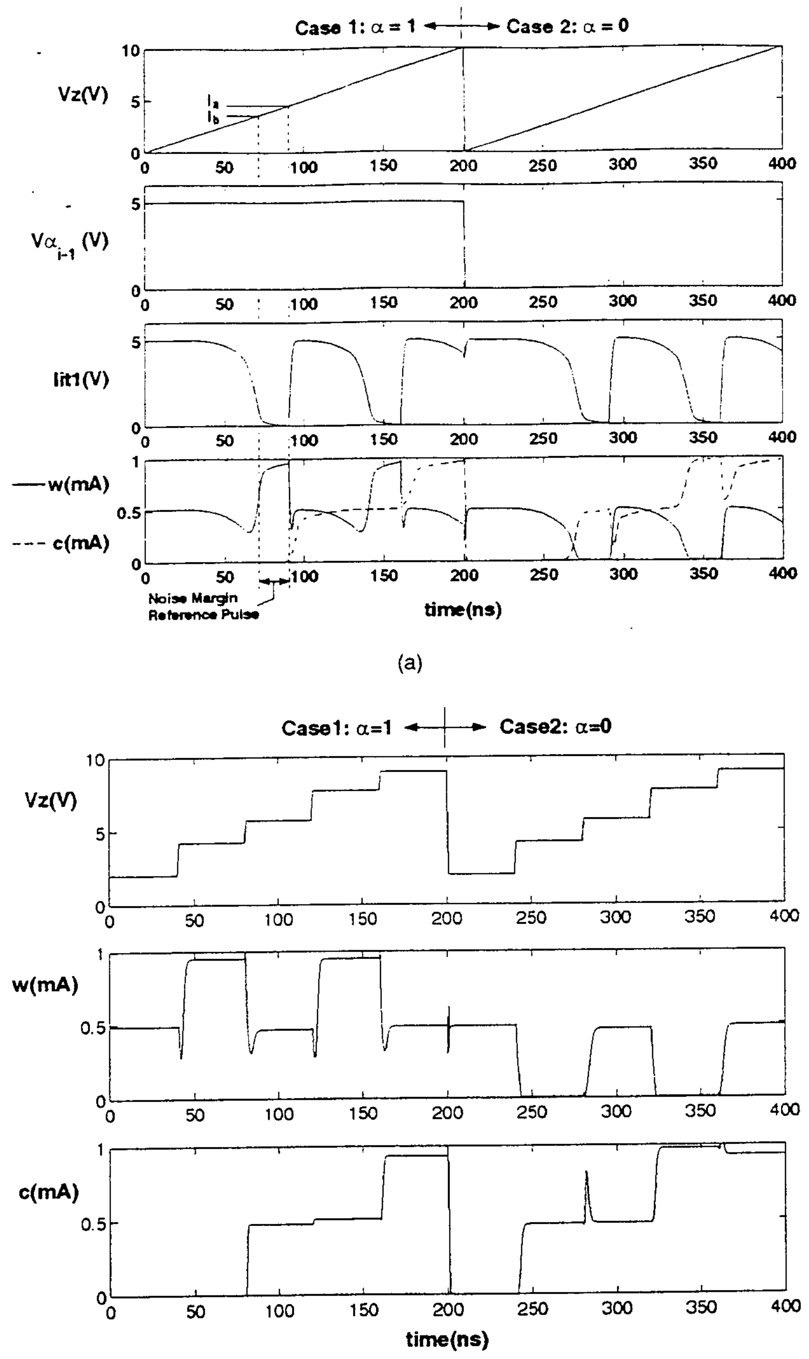
$$\alpha_1 = \frac{R_{p1} + R_s}{R_{p1} R_s\left(C_g + C_{be1}\right)}.$$

To find the initial conditions for the second part of the analysis, (5) is used to solve

$$V_z(t_a) - V_{node}(t_a) = V_p,$$

where t_a is the time at which the RTD enters the NDR region of operation. Using t_a, the final value, V_{za}, of the input signal when the operation of the circuit leaves the PDR region ($V_{za} = S_R t_a$) is determined. V_{za} is the initial value for the input variable V_z in the second part of the analysis.

The second part of the analysis is performed in a similar fashion. In order to make things more simple, a new time reference is used. Considering this assumption and the previous result, the input variable is expressed as



(a)

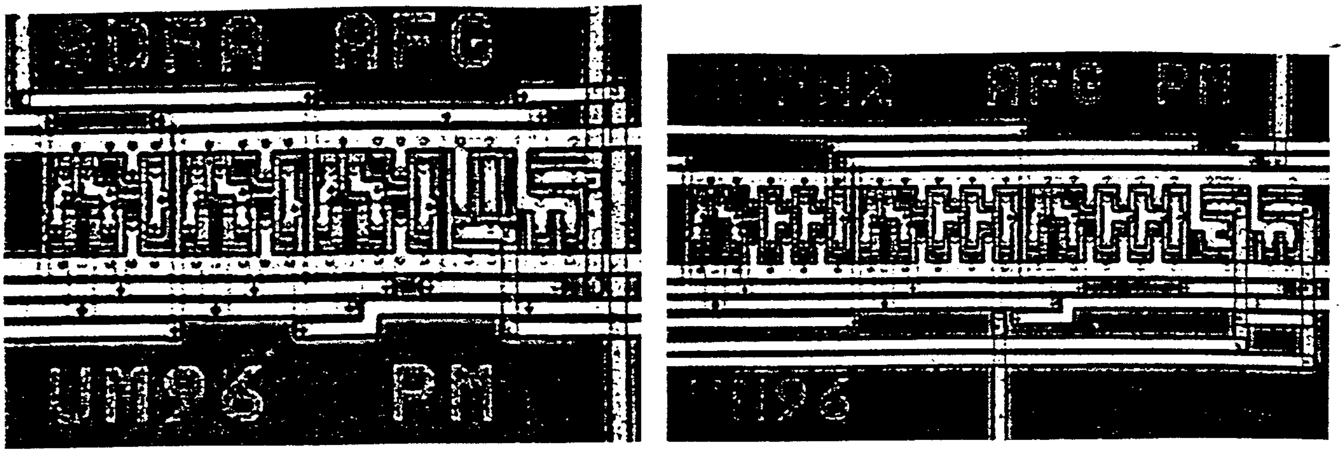
(b)

Fig. 11. Circuit simulation traces for the (a) first and the (b) second experiments.

pulse in output signal w (as shown in Fig. 11a). The levels i_a and i_b of input signal V_z at 50 percent of each transition of the selected pulse of w are measured. The noise margin is obtained by assuming that the operating point is at the middle of the selected output pulse. Hence, the noise margin is half of the difference between the measured current values of signal V_z, that is, $NM = |i_a - i_b|/2$. The value of power dissipation given in Table 2 was obtained in the second simulation experiment (Fig. 11b). The frequency of operation affects the value of the measured power dissipation. Finally, the delay of the circuit was measured as the time elapsed between 50 percent of the transition in the input signal, V_z, and 50 percent of the resulting transition in the output signal (in the second experiment). Both rising and falling delays were measured, but only the worst case (rising) result is given. A good characteristic of the proposed approach is that it will be able to take advantage of the progress in CMOS technology. For instance, the delay values will decrease as the circuit is implemented with more advanced CMOS processes.

4.2 Prototype Implementation

It was important to implement a working prototype in order to demonstrate the proposed principle of operation.



(a) (b)

Fig. 13. Microphotographs of the fabricated test chip.

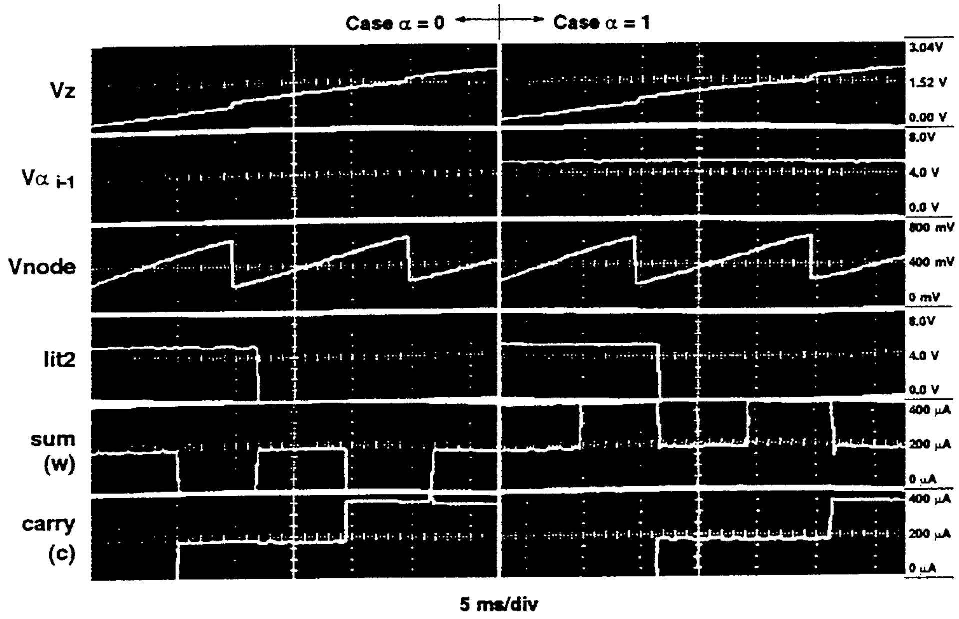


Fig. 14 Oscilloscope traces showing the operation of the signed-digit adder test circuit at low frequency.

values. While the expected output currents for logic levels "−1," "0," and "1" are 0.0, 0.5, and 1.0 milliampere, respectively, the measured current levels are 0.0, 0.2, and 0.4 milliampere (Fig. 14). This alteration of the current levels is due, in part, to a reduction in the transconductance of the PMOS devices—from 20.563 $\mu A/V^2$ in the SPICE model used for the simulations, to 16.129 $\mu A/V^2$ measured for the actual CMOS run of the test chips. Another reason for the difference in output levels is due to the method used for measuring the output currents. In the experiment, a resistive load was connected to the outputs in order to convert the current signals into voltage signals that could be displayed by the oscilloscope.

5 CONCLUSIONS

This paper presents a new multiple-valued signed-digit adder circuit that combines, for the first time, resonant-tunneling diodes (RTDs) with MOS field-effect transistors. RTDs provide high functionality for compact multiple-valued logic implementation, and MOS transistors enable efficient arithmetic circuit design through current-mode operation. The principle of operation of the proposed circuit was demonstrated using circuit simulation and also through a prototype fabrication where the GaAs-based quantum tunneling devices were not cointegrated with monolithically fabricated MOSFET devices. RTDs were externally added to verify the correctness of the SDFA operation as predicted by the simulation model.

The main advantage of the proposed design when compared to other redundant adder implementations is compactness, which is primarily due to the nonlinear characteristics of RTDs that enabled us to obtain three literal functions very efficiently. Also, current-mode of circuit operation, in which digits are summed by merely connecting their wires together [28], enabled us to reduce the transistor count. The number of devices is used as an estimate of the size of the circuit. This criterion is based on the fact that

[15] K.C. Smith, "The Prospects for Multivalued Logic: A Technology and Application View," *IEEE Trans. Computers*, vol. 30, no. 9, pp. 619-634, Sept. 1981.

[16] A.C. Seabaugh, Y.-C. Kao, and H.-T. Yuan, "Nine-State Resonant Tunneling Diode Memory," *IEEE J. Solid State Circuits*, vol. 13, pp. 479-481, Sept. 1992.

[17] H.C. Lin, "Resonant Tunneling Diodes for Multi-Valued Digital Applications," *Proc. Int'l Symp. Multiple-Valued Logic*, pp. 188-195, 1994.

[18] T. Waho, "Resonant Tunneling Transistor and Its Application to Multiple-Valued Logic Circuits," *Proc. Int'l Symp. Multiple-Valued Logic*, pp. 130-138, 1995.

[19] L.J. Micheel and H.L. Hartnagel, "Intraband RTDs with nanoelectronic HBT-LED Structures for Multiple-Valued Logic Computation," *Proc. Int'l Symp. Multiple-Valued Logic*, pp. 80-85, 1996.

[20] S. Kawahito, M. Kameyama, and T. Higuchi, "Multiple-Valued Radix-2 Signed-Digit Arithmetic Circuits for High-Performance VLSI Systems," *IEEE J. Solid State Circuits*, vol. 25, pp. 125-131, Feb. 1990.

[21] U. König, M. Kuisl, J.-F. Luy, and F. Schäffler, "Si/SiGe Resonant Tunneling Devices Separated by Surrounding Polysilicon," *Electronic Letters*, vol. 25, pp. 1,169-1,171, 1989.

[22] U. König, M. Kuisl, F. Schäffler, G. Fischer, and T. Kiss, "Operating CMOS After a Si-MBE Process: A Precondition for Future Three-Dimensional Circuits," *IEEE Electron Device Letters*, vol. 11, pp. 218-220, May 1990.

[23] N. Evers, O. Vendier, C. Chun, M.R. Murti, J. Laskar, N.M. Jokerst, T.S. Moise, and Y.-C. Kao, "Thin Film Pseudomorphic AlAs/InGaAs/InAs Resonant Tunneling Diodes Integrated onto Si Substates," *IEEE Electron Device Letters*, vol. 17, pp. 443-445, Sept. 1996.

[24] S.-J. Wei, H.C. Lin, R.C. Potter, and D. Shupe, "Dynamic Hysteresis of the RTD Folding Circuit and Its Limitation on the A/D Converter," *IEEE Trans. Circuits and Systems—II: Analog and Digital Signal Processing*, vol. 39, pp. 247-251, Apr. 1992.

[25] S. Mohan, J.P. Sun, P. Mazumder, and G.I. Haddad, "Device and Circuit Simulation of Quantum Electronic Devices," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, pp. 653-662, June 1995.

[26] S. Mohan, P. Mazumder, and G.I. Haddad, "NDR SPICE: A Circuit Simulator for Resonant Tunneling Devices," *Proc. Int'l Compound Semiconductors Conf.*, Sept. 1994.

[27] S. Mohan, P. Mazumder, and G.I. Haddad, "A New Circuit Simulator for Negative Resistance Devices," *Proc. Int'l Electron Devices Meeting*, Dec. 1994.

[28] M. Kameyama and T. Higuchi, "Design of a Radix-4 Signed-Digit Arithmetic Circuit for Digital Filtering," *Proc. Int'l Symp. Multiple-Valued Logic*, pp. 272-277, 1980.

Alejandro F. González (S'96) received the BE degree in electrical engineering (Licenciado en Ingenierá Electrónica) from the Instituto Tecnológico y de Estudios Superiores de Occidente, Guadalajara, Mexico, in 1993, and the MSE degree in electrical engineering from the University of Michigan, Ann Arbor, in 1995. He is currently pursuing the PhD degree in electrical engineering at the University of Michigan. His research interests include ultrafast digital circuit design, multiple-valued logic, and VLSI design automation. He is a student member of the IEEE.

Pinaki Mazumder (S'84–M'87–SM'95) received the BSEE degree from the Indian Institute of Science in 1976, the MSc degree in computer science from the University of Alberta, Canada, in 1985, and the PhD degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1987.

For two years, he was a research assistant with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. For more than six years, he was with Bharat Electronics Ltd. (a collaborator of RCA), India, where he developed analog and digital integrated circuits for consumer electronics products. During the summers of 1985 and 1986, he was a member of the technical staff in the Naperville, Illinois, branch of AT&T Bell Laboratories. He spent his sabbatical year as a visiting faculty member at Stanford University, University of California, Berkeley, and also at Nippon Telegraph and Telephone (NTT), Japan. He is presently with the Department of Electrical Engineering and Computer Science of the University of Michigan, Ann Arbor. His research interests include VLSI testing, physical design automation, and ultrafast digital circuit design. He has written more than 100 papers for archival journals and international conference proceedings on these topics. He has coauthored two books entitled *Testing and Testable Design of Random Access Memories* (Kluwer, 1996) and *Genetic Algorithms for VLSI Layout and Test Automation* (Prentice Hall, 1998).

Dr. Mazumder was a recipient of Digital's Incentives for Excellence Award, U.S. National Science Foundation Research Initiation Award, and Bell Northern Research Laboratory Faculty Award. He is an associate editor of *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. He was a guest editor of *IEEE Design and Test*s special issue on multimegabit memory testing, March 1993, a guest editor of *Journal of Electronic Testing: Theory and Applications*' special issue on advanced techniques for memory testing, April 1994, and a guest editor of *IEEE Transactions on VLSI Systems*' special issue on the impact of emerging technologies in VLSI systems, March 1998. He is a member of Sigma Xi, Phi Kappa Phi, the ACM Special Interest Group on Design Automation, and a senior member of the IEEE.

Full Adder Circuit Design Using RTDs and MOSFETs

Shriram Kulkarni and Pinaki Mazumder
Department of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109

Abstract

This paper presents the design of a bistable-mode full adder circuit based on a new family of logic gates using resonant tunneling diodes (RTDs) and n-MOSFETs. The high switching speed of RTDs and their negative differential-resistance (NDR) characteristics are ideally suited for the design of fast and compact self-latching logic circuits. The possible use, in the future, of these bistable gates in deeply pipelined computing systems will result in elimination of the area and delay overhead of pipeline latches that limits conventional CMOS-based systems, leading to compact and fast system designs while retaining advantages of CMOS, such as low power dissipation and high packing density.

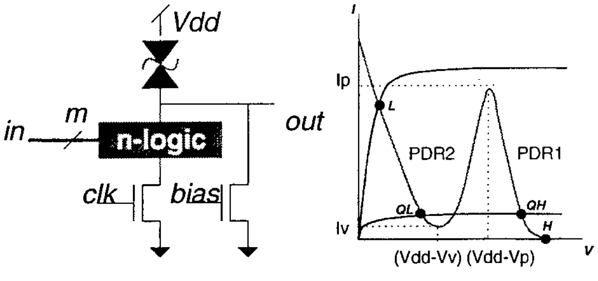
1 Introduction

Several novel memory and logic circuits using resonant tunneling diodes (RTDs) [1, 2, 3] and RTDs in conjunction with transistors [4, 5, 6] have been proposed in the past few years . These circuits have one or more of the following advantages over conventional logic circuits: 1) reduced circuit complexity for implementing a given function, 2) low power operation, and 3) high speed operation. RTD-based circuits are usually fabricated using III-V technology which offers high performance devices such as hetero-junction bipolar transistors (HBTs) and high electron mobility transistors (HEMTs). While such circuits demonstrate very high switching speeds, the operate at high current levels thus consuming substantially higher absolute power than equivalent CMOS circuits. The packing density of III-V technology based VLSI circuits is more than an order of magnitude smaller than CMOS VLSI circuits. However, the unique NDR characteristics of RTDs coupled with their high tunneling speeds lead to very compact and fast circuit topologies. Thus it is very attractive to envision these compact, high-functionality circuits implemented in a technology such as CMOS that offers very low power dissipation and very large integration levels. While co-integration of RTDs and MOSFETs is still an active area of research, in the following sections, we discuss the operating principle of a self-latching RTD-MOSFET digital logic gate and the design of a pipelined full adder circuit using RTDs and MOSFETs.

2 RTD-MOSFET Bistable Logic

A binary logic circuit is said to operate in bistable mode when its output is latched, and any change in the input is reflected in the output only when a clock or other evaluation signal is applied. The bistable mode has been used in several earlier technologies, notably in superconducting logic [7]. Superconducting logic typically uses a multi-phase AC power source to periodically reset/evaluate each gate. Similar logic using resonant-tunneling devices has been proposed by other authors [5, 8]. The chief disadvantage of these circuits is the requirement of an AC power source whose frequency determines the maximum switching frequency. The RTD-MOSFET logic circuits described herein use a DC power supply and multiphase clocks but the clock signals are not required to supply large amounts of power as in the case of the earlier circuits. The operating principle of the new bistable element may be understood by considering the simplified circuit shown in Figure 1a. The RTD forms the active load in the circuit. There are m inputs in the n-MOSFET logic block which determines the circuit function, a clock transistor which controls the evaluation of the gate, and a bias transistor that maintains the quiescent current through the RTD while also controlling the precharging of the gate output. Figure 1b shows the load lines for the bistable logic gate. The quiescent current maintained by the bias transistor is between the peak and valley currents of the RTD. The clock transistor is such that if the n-logic block is turned on, the arrival of the clock pulse will cause the current through operation of the RTD-MOS gate is enumerated below.

- Inputs I_1 through I_m change. The clk signal is held low thus preventing evaluation of the output node by the n-logic block.
- The $bias$ signal goes low forcing the current through the RTD to zero. When the current falls below the valley current, I_v, of the RTD, the out node is pulled high. This is point H as indicated in Figure 1b.
- The $bias$ signal is then set to logic high. The out node remains high. The current through the RTD is now the quiescent current, I_q, and the circuit operating point changes to QH in the load lines of Figure 1b.
- The clk signal goes high. If the inputs, I_1 through I_m are such that the n-logic block is turned on, then turning the clock transistor on causes the current through the RTD to exceed the peak current, I_p, of the RTD causing a jump to the second positive differential resistance ($PDR2$) region of the RTD characteristic corresponding to $V_{RTD} > V_v$, where V_{RTD} is the voltage across the RTD and V_v is the valley voltage of the RTD. This results in the out node going low. The circuit is at operating point L. If the n-logic block is not turned on, the RTD current does not increase beyond I_q and the circuit operating point remains at QH in the first positive differential resistance region ($PDR1$) of the RTD, where $V_{RTD} < V_p$, V_p being the RTD peak voltage. Thus, out remains high.
- The clk signal goes low so that no current flows through the n-logic block. The output voltage at node out reaches a stable value corresponding to whether the RTD was in $PDR1$ (operating point QH) or $PDR2$ (operating point QL) in the previous step of the sequence.



(a) (b)

Figure 1: (a) Bistable-mode RTD-MOSFET logic gate (b) Load lines for RTD-MOSFET gate operating point

The operating principle of the RTD-MOSFET gate presents some design constraints for the sizes of the RTD and n-MOSFETs used in the circuit. Let m be the area of the RTD used and let J_p and J_v represent the peak and valley current densities of the RTD, respectively. Let β_n represent the effective gain factor of the n-logic block and the clock transistor. Let β_{bias} be the gain factor of the bias transistor. Let I_q be the quiescent operating current and let I_h be the ON current flowing through the n-logic block on the arrival of the clock pulse. Also, if we assume that the bias transistor is in saturation at operating points QH and QL, and the n-logic block and clock transistor are in saturation when the gate is positively evaluated in $PDR1$, then the design constraints for the aforementioned gate can be written as:

$$
\begin{aligned}
mJ_p - I_q &> 0, && (1)\\
I_q - mJ_v &> 0, && (2)\\
I_h + I_q - mJ_p &> 0, && (3)\\
I_h, I_q &> 0. && (4)
\end{aligned}
$$

where,

$$
\begin{aligned}
I_q &= \beta_{bias}(V_{dd} - V_{tn})^2/2, && (5)\\
I_h &= \beta_n(V_{dd} - V_{tn})^2/2. && (6)
\end{aligned}
$$

To implement a bistable 2-input NAND gate, we would require 4 n-MOS transistors and one RTD. In comparison, a static CMOS 2-input bistable NAND gate would require 4 transistors for the NAND logic and 10 transistors for an output latch. This represents a substantial reduction in circuit size for RTD-MOSFET logic gates as compared to static CMOS which translates to smaller parasitic capacitances and in conjunction with the high switching speed of RTDs, leads to substantial improvement in circuit speed. In light of the bistable nature of the RTD-MOSFET gates, it is more relevant to compare these new gates with dynamic CMOS logic gates. Table 1 shows the comparison of simulated values of power and delay for a bistable 2-input NAND gate designed using dynamic CMOS logic and RTD-MOSFET logic. Both designs use a supply voltage of 2V, assume a $1\mu m$ CMOS process and use identical pull-down networks for the n-logic and clock.

3 Full Adder With Gate-level Pipelining

Using the RTD-MOSFET logic, a bistable full adder is designed as a two-stage logic block. For correct

Table 1: Comparison of bistable $NAND2$: Dynamic CMOS vs. RTD-MOSFET

Parameter	Dynamic CMOS	RTD-MOSFET
Device Count	5	5
Power (mW)	0.47	0.52
$(t_r + t_f)$(ns)	1.42	0.46
Power-Delay (pJ)	0.67	0.24

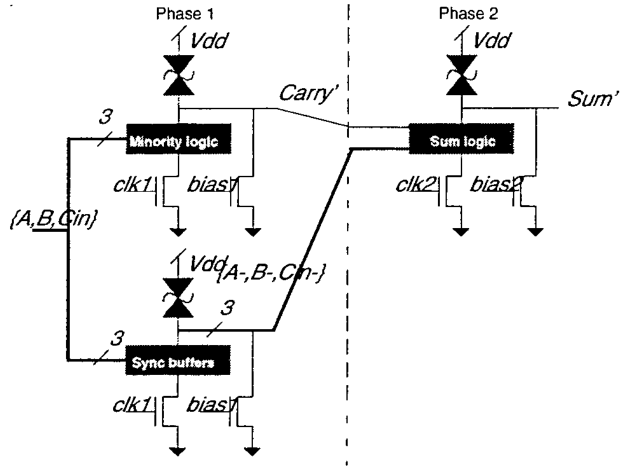


Figure 2: Pipelined RTD-MOSFET full adder schematic

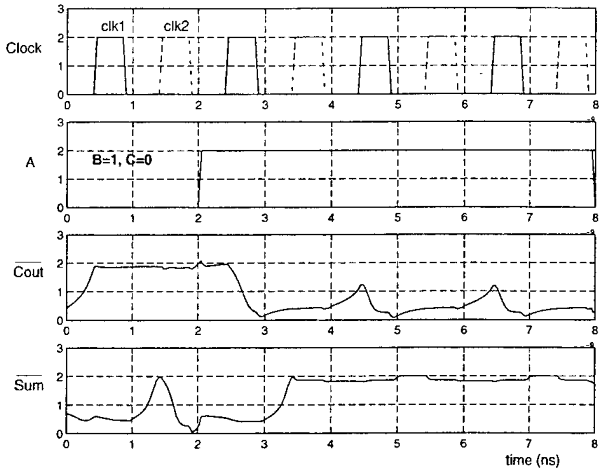


Figure 3: Pipelined RTD-MOSFET full adder simulation

operation of the bistable logic gates, bias and clock pulses are required as mentioned previously. However, when multiple gates are cascaded, as is the case in a full adder, a gate must be clocked only after all its inputs have been correctly evaluated. This requires a multiphase clocking scheme in which each gate is evaluated in a different phase than its fanins and fanouts. For cascaded RTD-MOSFET gates, a two-phase clocking scheme is used. This is similar to dynamic CMOS clocking. Figure 2 shows the schematic of the full adder circuit. The majority function that gives the carry output is evaluated on $Phase1$ of the clock. Each clock phase consists of a pulse on the bias line followed by an evaluation clock pulse in accordance with the operation sequence of the gate explained in Section 2. Bistable buffers clocked on $Phase1$ are required to synchronize inputs to the sum stage of the circuit which is evaluated on $Phase2$ of the clock. Simulation traces of the 1-bit full adder circuit using nMOS pulldown logic in $1\mu m$ CMOS process technology are shown in Figure 3. The rise and fall times of the Sum signal are 0.4 ns each. The RTD-MOSFET adder is pipelined at the gate level without necessity for external pipeline latches. This has advantages over conventional pipelining schemes that are commonly used to speed up computation by dividing a combinational block into several sequential stages such that each stage performs a different operation during a particular clock cycle. Conventional pipelining schemes are limited by the added delay and area of pipeline latches. A combinational block consisting of n stages, each having a delay of t_c, will have a total delay of nt_c. If we pipeline this circuit by introducing a latch, having delay t_l, at the output of each stage, the maximum delay of the circuit becomes $(t_c + t_l)$. The throughput of the circuit increases from $1/(nt_c)$ to $1/(t_c + t_l)$ but the latency increases from nt_c to $n(t_c + t_l)$. Also, if a_c is the area of the combinational block; after pipelining, the area of the circuit increases to $a_c + nka_l$, where a_l is the area of a latch and k is the number latches at each stage. If the latch delay t_l is much larger than the stage delay t_c, it places an upper bound on the maximum achievable throughput of the pipelined circuit. Thus, in conventional pipelining schemes, there exist direct tradeoffs between the the area of the pipeline latches and the achievable throughput. However, the use of RTDs and MOSFETs in designing self-latching circuits improves the performance of these gate-level pipelined circuits conventional pipelined circuits because the latch delay, $t_l = 0$. Also, if latency is not of concern, as is the case in signal processing systems and communication systems, each logic gate can operate in the bistable mode resulting in maximum possible throughput.

The RTD-MOSFET bistable full adder uses 5 RTDs and 20 n-MOSFETs. To convert a standard 24-MOSFET static CMOS adder to a similar gate-level pipelined adder, we would require an addiresulting in maximum possible throughput.

The RTD-MOSFET bistable full adder uses 5 RTDs and 20 n-MOSFETs. To convert a standard 24-MOSFET static CMOS adder to a similar gate-level pipelined adder, we would require an additional 40 transistors for 5 latches required for the carry, the sum and the three input signals. The addition of these latches would provide the pipelining advantage in standard CMOS but would increase the stage delay and area due to the additional latches. Thus, the RTD-MOSFET logic family has advantage over conventional CMOS logic in terms of greater circuit compactness and improved speed. The bistable nature of these circuits eliminates pipeline latch overhead of area and delay, leading to higher throughput in pipelined systems.

3.1 Design and Implementation Issues

Co-integration of RTDs and MOSFETs is still an active area of research, and at the present time it is not possible to fabricate RTDs in silicon that have usable NDR characteristics. However, since the feasibility of the RTD-MOSFET logic family has been demonstrated, improvements in process technology should make these circuits viable in high-speed switching applications. Switching margins and power dissipation are conflicting constraints that should be optimized during the logic design of each gate. As in any VLSI design, clock and timing considerations are paramount due to the use of level sensitive, non-overlapping two-phase clocks. With the use of an output buffer, fanout of the gates may be considerably improved while isolating the RTD-MOSFET node from loading effects. Also, the use of an output buffer allows the use of the proposed logic family in the domino mode requiring only a single-phase clock for a cascaded logic structure.

4 Conclusions

A design methodology for a new logic family using RTDs and MOSFETs has been presented. The proposed logic family has advantage over conventional CMOS logic in terms of circuit compactness and improved power-delay product. Basic bistable logic gates have been designed in the proposed logic family and their operation has been verified through simulation. The design of complex functions such as a pipelined full adder has been demonstrated. The bistable nature of the circuits eliminates pipeline latch overhead of area and delay, leading to higher throughput in pipelined systems. In light of this, the proposed logic family has possible applications in high speed communication systems and signal processing systems where a large fraction of the devices are switching at a given time.

References

[1] F. Capasso, S. Sen, F. Beltram, L. M. Lunardi, A. S. Vengurlekar, P. R. Smith, N. J. Shah, R. J. Malik, and A. Y. Cho, "Quantum functional devices: resonant-tunneling transistors, circuits with reduced complexity and multiple-valued logic," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2065–2082, October 1989.

[2] M.-H. Shieh and H. C. Lin, "A multiple-dimensional multiple-state sram cell using resonant tunneling diodes," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 623–630, May 1994.

[3] W. Williamson III, S. Enquist, D. Chow, H. Dunlap, S. Subramaniam, P. Lei, G. Bernstein, and B. Gilbert, "12 GHz clocked operation of ultralow power interband resonant tunneling diode pipelined logic gates," *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 222–231, February 1997.

[4] A. C. Seabaugh, A. H. Taddiken, E. A. Beam III, J. N. Randall, Y. C. Kao, and B. Newell, "Co-integrated resonant tunneling diode and heterojunction bipolar transistor full adder," in *International Electron Devices Meeting*, pp. 419–422, 1993.

[5] K. Maezawa and T. Mizutani, "A new resonant tunneling logic gate employing monostable-bistable transition," *Japanese Journal of Applied Physics*, vol. 32, pp. L42–L44, January 1993.

[6] H. Chan, S. Mohan, P. Mazumder, and G. Haddad, "Compact multiple-valued multiplexers using negative differential resistance devices," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 1151–1156, August 1996.

[7] T. Van Duzer, "Superconductor digital ICs," in *VLSI Handbook* (J. DiGiacomo, ed.), New York: McGraw-Hill, 1989.

[8] S. Mohan, P. Mazumder, R. K. Mains, J. P. Sun, and G. I. Haddad, "Logic design based on negative differential resistance characteristics of quantum electronic devices," *IEE Proceedings-G*, vol. 140, pp. 383–391, December 1993.

TABLE II
Results for rise and fall times

		Analytical	Simulation
Ex 1	t_{rise}(ps)	17.82	15.69
	t_{fall}(ps)	49.95	53.78
Ex 2	t_{rise}(ps)	25.78	22.16
	t_{fall}(ps)	60.67	63.41
Ex 3	t_{rise}(ps)	32.44	27.47
	t_{fall}(ps)	62.05	74.57

$$\begin{aligned} \text{where } A^2 &= [1-\beta|R_N|(V_{DD}-V_{TN0})]^2 \\ &\quad -2\beta|R_N|[I_P|R_N|+V_P-V_{DD}] \\ \text{and} \quad B &= \beta|R_N|(V_V-V_{TN0}), \end{aligned}$$

$$\begin{aligned} t_{f_{P_2}} &= \left[\tfrac{C_L R_{P_2}}{A'}\right] \ln \left[\tfrac{\beta R_{P_2}(V_{OL_1}-V_{DD}+V_{TN0})-1-A'}{\beta R_{P_2}(V_{OL_1}-V_{DD}+V_{TN0})-1+A'} \right. \\ &\quad \left. \cdot \tfrac{B'+1-A'}{B'+1+A'}\right], \end{aligned} \tag{4}$$

$$\begin{aligned} \text{where } A'^2 &= [1+\beta R_{P_2}(V_{DD}-V_{TN0})]^2 \\ &\quad -2\beta R_{P_2}[I_V R_{P_2}+V_{DD}-V_V] \\ \text{and} \quad B' &= \beta R_{P_2}(V_V-V_{TN0}). \end{aligned}$$

The rise time (t_{rise}) is similarly divided into three parts, which are:

$$t_{r_{P_2}} = R_{P_2} C_L \ln \left[\frac{I_V R_{P_2} + (V_{DD} - V_{OL_2}) - V_V}{I_V R_{P_2}}\right], \tag{5}$$

$$t_{r_N} = |R_N| C_L \ln \left[\frac{I_P}{I_V}\right] \tag{6}$$

$$\text{and} \quad t_{r_{P_1}} = R_{P_1} C_L \ln \left[\frac{V_P}{V_{DD} - V_{OH}}\right]. \tag{7}$$

Using typical 0.5μm CMOS process parameters and a realistic RTD characteristic, we compared the results obtained from the above equations with SPICE simulation results for a few different transistor sizes. The results are presented in Table II.

Expressions for power consumption

Consider the first inverter shown in Fig.1. The power consumption of the inverter is given by

$$P_d = \frac{1}{T}\int_0^T I_r(t) V_{DD}\, dt. \tag{8}$$

When IN is low, the power consumption is given by:

$$P_d = \int_0^T V_{DD} I_r(t) dt = \frac{t_{eval} + t_{latch}}{T} V_{DD} \cdot I_{OL2}. \tag{9}$$

The current waveform when IN is high is shown in Fig. 3. As shown in the figure, we can divide one period into five regions and derive analytical expressions for the energy spent in these time ranges. The corresponding energy expressions are:

$$E_{t_{f1}} = \frac{1}{2} V_{DD} \cdot I_P \cdot t_{f1}, \tag{10}$$

$$E_{t_{f2}} = \frac{1}{2} V_{DD} \cdot (I_P + I_V) \cdot t_{f2}, \tag{11}$$

$$E_{t3} = V_{DD}(I_{OL1} t_3 - (I_{OL1} - I_V)\tau_{f3}(1 - e^{-\frac{t_3}{\tau_{f3}}})), \tag{12}$$

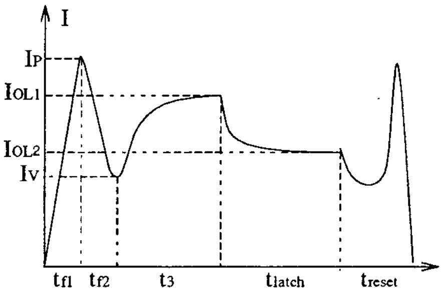


Fig. 3. Current waveform of RTD when IN is high

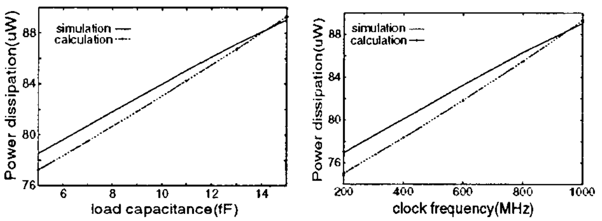


Fig. 4. Dependence of power consumption on load capacitance and frequency

$$E_{latch} = V_{DD} I_{OL2} t_{latch} + V_{DD}(I_{OL1} - I_{OL2})\tau_r(1 - e^{-\frac{t_{latch}}{\tau_r}}), \tag{13}$$

$$E_{reset} = V_{DD} \int_{V_{OL2}}^{V_{DD}} C_L dV_O = V_{DD} C_L (V_{DD} - V_{OL2}). \tag{14}$$

Hence, the total power consumption when IN is high is given by:

$$P_{d(IN=high)} = \frac{1}{T}(E_{t_{f1}} + E_{t_{f2}} + E_{t_3} + E_{latch} + E_{reset}). \tag{15}$$

The average power consumption is given by

$$P_d = \frac{1}{2}(P_{d(IN=low)} + P_{d(IN=high)}). \tag{16}$$

A comparison of our analytical result with SPICE simulation of bistable QMOS inverter's power consumption is presented in Fig. 4. Both simulation and calculation indicates that power consumption of QMOS inverter doesn't increase much with frequency and load capacitance.

Conclusions

In this paper, we have presented analytical expressions for delay and power consumption of bistable QMOS inverter circuit. Our analysis shows that 60–70 ps gate delays with 80–90 μW of power consumption can be expected from such gates.

References

[1] N. Evers, O. Vendier, C. Chun, M. R. Murti, J. Laskar, N. M. Jokerst, T. S. Moise, and Y.-C. Kao, "Thin Film Pseudomorphic $AlAs/In_{0.53}Ga_{0.47}As/InAs$ Resonant Tunneling Diodes Integrated onto Si Substrates," *IEEE Electron Device Letters*, 1996.

[2] J. Xiong, N. Deb, M. Bhattacharya, S. Kulkarni, and P. Mazumder, "Analysis of bistable qmos circuits." under preparation.

Appendix 52

Ultra-fast Adder Circuit Design using Resonant Tunneling Diodes and MOSFETS

Alejandro F. González and Pinaki Mazumder
Electrical Engineering and Computer Science Department
The University of Michigan, Ann Arbor, MI 48109

Abstract

This paper presents a novel signed-digit adder design using current mode logic and resonant-tunneling diodes (RTDs). A current mode logic, which allows wired addition of signals, is implemented using MOS transistors. The circuit uses the folded I-V characteristics of the RTD to achieve a very compact implementation of the multivalued adder functions.

1 Introduction

Negative differential-resistance (NDR) devices, such as RTDs, can be used to implement many analog and digital circuit applications by exploiting their folded I-V characteristics to achieve very compact circuit designs [1]. Although the properties of RTDs were observed more than two decades ago [2], it was not until epitaxial growth techniques were developed that the fabrication of devices with large, hence useful, NDR effects was made possible.

Signed-digit number representations were proposed in 1961 by Algirdas Avizienis [3]. This type of representations eliminate carry propagation chains by limiting the transmission of carry signals to one digit position during addition and subtraction. Signed-digit representations have an inherent redundancy that allows an addition algorithm where each digit of the resulting sum is function of only two input operand digits.

This paper presents a signed-digit adder design based on a combination of RTD and CMOS devices. The proposed design uses *current-mode* signal representation which allows *wired addition* of signals [4]. The adder implements radix-2, signed-digit addition, and hence requires a three-level logic (using digits $\{-1, 0, +1\}$). This new signed-digit adder uses 13 MOS transistors, while 44 transistors are used in an equivalent redundant-binary static CMOS implementation.

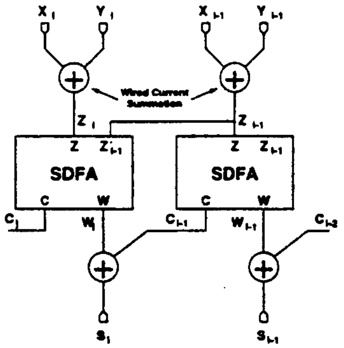


Figure 1: Block diagram of the addition approach.

2 Adder Output Functions

Figure 1 depicts a block diagram of the signed-digit addition approach proposed in this work. Lines x_i, y_i, c_i, w_i, and s_i are three-valued, current-mode signals. Addition of x_i and y_i is performed by simple wired-summation of currents. The function of the SDFA block is to convert the summation input signal, z, to a two-digit representation of the sum given by digits c and w ($rc + w = z$, where $r = 2$). The final sum output, s_i, is obtained by current-addition of the interim sum output, w_i, and the incoming carry signal, c_{i-1}.

Figure 2 shows the transfer functions for the interim sum, w, and the carry, c, signals in the SDFA cell. All the digits in the graph are positive because the circuit uses only positive currents. In this case, the signed-digit 0 is represented by a current level "3", digit -2 is represented by current "1", and so on. Note that there are two pairs of transfer functions and the working pair is selected by the value of z_{i-1}. This input signal is used to determine the condition $c_{i-1} \neq -1$ so that, when required, the SDFA cell is able to generate an output $w = -1$ and still keep a valid three-level sum output.

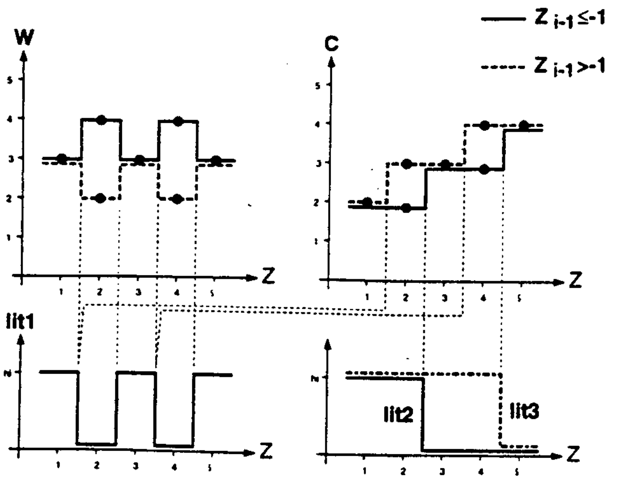


Figure 2: Transfer function of the SDFA block.

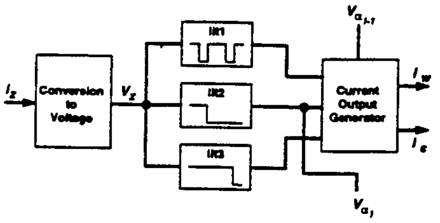


Figure 3: Block diagram of the SDFA cell.

3 Adder Design

Careful observation of the SDFA transfer characteristics indicates that three literal signals can be used to describe the multivalued functions. As shown in Figure 2, literals $lit1$, $lit2$, and $lit3$ contain all the switching information required to define output functions w and c. This observation comprises the basis of the proposed SDFA cell, whose block diagram is shown in Figure 3. The input to the system is the wired addition current signal i_z. The three required literal signals are then generated in different blocks of the circuit. The literal signals are used to control switched current sources, which in turn synthesize the SDFA transfer functions in the *current output generator* block. The behavior of output V_{α_i} is identical to that of literal $lit2$. Note that the output block uses input signal $V_{\alpha_{i-1}}$ to determine which of the two sets of transfer characteristics should be used. This input signal indicates if $z_{i-1} < -1$.

3.1 Literal Circuit Implementation

Figure 4 shows the basic circuit used for generating literal $lit1$. The circuit consists of a series connection of the resistor R_s and the RTD, with the latter performing as the load. The input to the circuit is the voltage V_z, which represents the summation $z = x + y$. Two RTDs are used in series to obtain an equivalent two-peak characteristic. In the circuits for generating $lit2$ and $lit3$, the RTD load is replaced by a simple resistor because $lit2$ and $lit3$ are simple threshold functions.

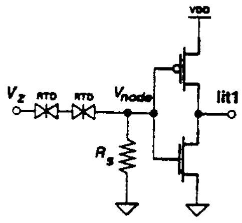


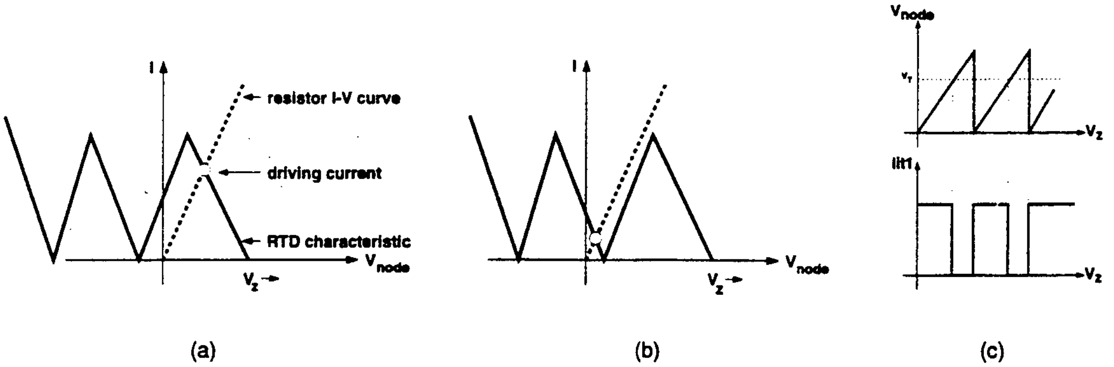
Figure 4: Compact circuit for generating $lit1$ using RTD and CMOS devices.

The behavior of the voltage V_{node} can be described using the load-line method (depicted in Figures 5(a) and 5(b)). In Figure 5, the I-V characteristics of the RTD and the resistor are plotted as the current I that circulates through the RTD/resistor versus the voltage V_{node}. The value of the current I is found at the point of intersection of the I-V curves. Note that the operating point of the circuit changes as the input voltage V_z is increased because the I-V curve of the RTD is moved to the right as V_z increases. Figure 5(c) depicts the formation of the literal signal $lit1$ by sensing V_{node} with the CMOS inverter.

3.2 Output Current Circuits

Figure 6(a) shows the current output generator for interim sum, w. In this circuit, transistor $m1$ injects a current equivalent to one logic level when $lit1$ is high. This states correspond to $z = 1, 3, 5$ in the transfer function shown in Figure 2. The operating points for $z = 2, 4$ are handled by transistors $m2$ and $m3$, which inject a current equivalent to two logic levels or no current at all, depending upon the value of $V_{\alpha_{i-1}}$.

The circuit for implementing function c is shown in Figure 6(b). In this circuit, transistors $m1$ and $m2$ produce the basic stairway form of the carry function for the case $z_{i-1} \leq -1$. Transistors $m3$ and $m4$ correct the current output c for $z_{i-1} > -1$ by injecting one logic level of current in the operating points where the two carry output functions are different ($z = 2, 4$). Note that $lit1$ controls the injection of the correcting current because this literal is low only in the operating points of interest ($z = 2, 4$).



(a) (b) (c)

Figure 5: (a), (b) Application of the load-line method in describing the behavior of the voltage V_{node}. (c) Ideal representation of the use of the CMOS inverter in the formation of $lit1$.

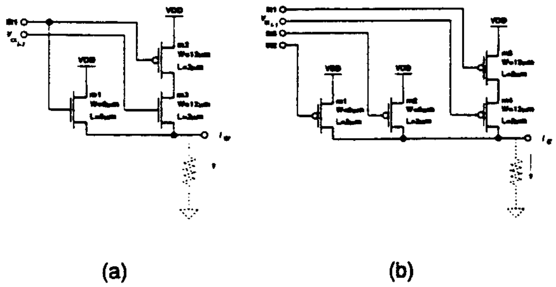


Figure 6: Circuits for generating the (a) interim sum and the (b) carry output functions.

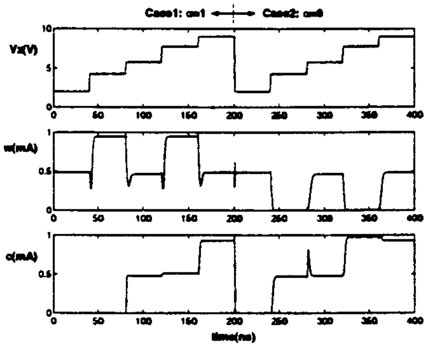


Figure 7: SDFA circuit simulation.

3.3 Simulation

The SDFA cell circuit was designed and verified using NDR-SPICE [5]. Figure 7 shows a transient analysis output trace obtained from a simulation of the SDFA circuit. This experiment includes both cases for input $V_{\alpha_{i-1}}$ ($z_{i-1} \leq -1$ and $z_{i-1} > -1$). Different simulation experiments were done to estimate performance measures such as delay, power dissipation and noise margins.

4 Experimental Results

4.1 Test Chip Implementation

We designed a small test circuit using a standard 2-micron CMOS process. This being an initial effort in the construction of a prototype, RTD and resistor elements were added as discrete external devices for testing the functionality of the design. The test chip includes circuits necessary for one adder cell, as well as independent instances of every building block (for improved testability).

At this point, the working implementation of the design does not stem directly from our previous calculations and simulations. The main problem is that the characteristics of the existing RTD do not meet the requirements of the design, given the threshold voltages of the CMOS process being used. The only choice is to make modifications in the design so that it becomes functional with the given devices.

The modification of the design consisted in replacing the inverter of the literal circuit shown in Figure 4 by a voltage comparator. Although this is not the most efficient solution, it is the most flexible. The use of a comparator is considered advantageous and appropriate for the test chip prototype. However, a mature implementation should aim at using the simple inverter. Figure 8 shows a microphotograph of the modified SDFA cell included in the chip.

The SDFA cell shown in Figure 8 is implemented with a voltage comparator that generates the com-

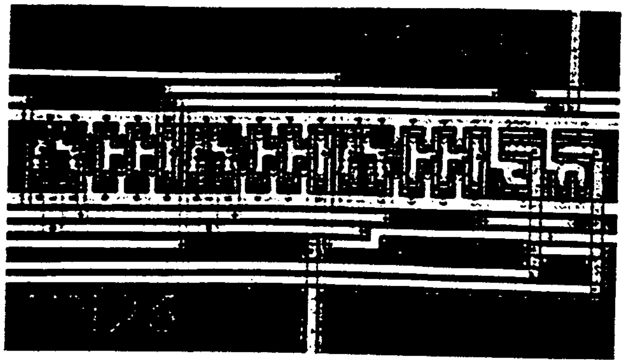


Figure 8: Microphotograph of the fabricated test chip. MOS circuitry for the modified SDFA cell.

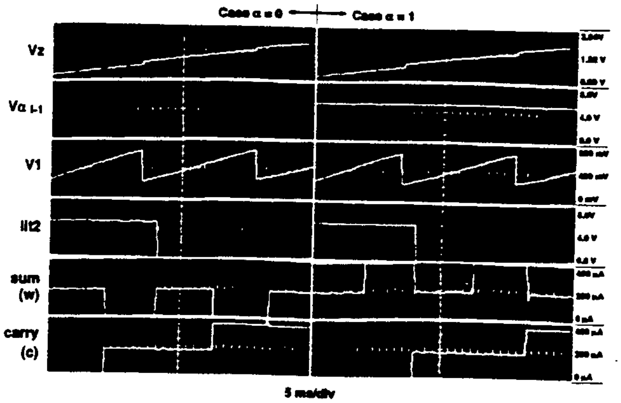


Figure 9: Oscilloscope traces showing the operation of the signed-digit adder test circuit at low frequency.

parison output as well as its complement. Generating the literals and their complements allows a circuit for the sum current output generation (w) which uses only pMOS transistors. The current generation circuit using only pMOS devices provides output signals with current levels of improved precision.

4.2 Functional Testing

The transfer characteristics were obtained by feeding the input of the circuit with a ramp signal of very low slew-rate. The experiment was done using the SDFA cell shown in Figure 8. The oscilloscope traces obtained in the experiment are shown in Figure 9. In general, the form of the measured output functions agree with the expected circuit behavior shown in Figure 7. Note that the experiment is divided into two parts with respect to the value of $\alpha = V_{\alpha_{i-1}}$. The trace labeled "$V_1$" reflects the operation of the RTD and corresponds to the expected behavior seen in Figure 5(c). Also, note that the behavior of literal $lit2$ is independent of the value of α. The last two traces show the transfer characteristics for the output functions, w and c, of the SDFA cell. Observe how the output characteristics are selected by α.

While the expected output currents for logic levels "-1", "0", and "1" are 0.0, 0.5, and 1.0 milliampere, respectively, the measured current levels are 0.0, 0.2, and 0.4 milliampere (Figure 9). This alteration of the current levels is due, in part, to a reduction in the transconductance of the pMOS devices—from 20.563 μA/V^2 in the SPICE model used for the simulations, to 16.129 μA/v^2 measured for the actual CMOS run of the test chips.

5 Conclusions

A new multiple-valued signed-digit adder cell has been presented. The proposed design achieves a compact implementation using only 13 transistors. A prototype was built for demonstrating the required transfer functions. The next step is to construct a prototype that allows the operation of the adder at very high speeds by integrating all the devices in a monolithic circuit.

References

[1] F. Capasso, S. Sen, F. Beltram, L. M. Lunardi, A. S. Vengurlekar, P. R. Smith, N. J. Shah, R. J. Malik, and A. Y. Cho, "Quantum functional devices: Resonant-tunneling transistors, circuits with reduced complexity, and multiple-valued logic," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2065–2082, October 1989.

[2] L. L. Chang, L. Esaki, and R. Tsu, "Resonant tunneling in semiconductor double barriers," *Applied Physics Letters*, vol. 24, pp. 593–595, 1974.

[3] A. Avizienis, "Signed-digit number representations for fast parallel arithmetic," *IRE Transactions on Electronic Computers*, pp. 389–400, September 1961.

[4] M. Kameyama and T. Higuchi, "Design of a radix-4 signed-digit arithmetic circuit for digital filtering," in *Proceedings of the International Symposium on Multiple-Valued Logic*, pp. 272–277, IEEE, 1980.

[5] S. Mohan, J. P. Sun, P. Mazumder, and G. I. Haddad, "Device and circuit simulation of quantum electronic devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, pp. 653–662, June 1995.

Compact Signed-Digit Adder Using Multiple-Valued Logic*

Alejandro F. González and Pinaki Mazumder

Electrical Engineering and Computer Science Department
The University of Michigan, Ann Arbor, Michigan 48109-2122
E-Mail: `{algonz,mazum}@eecs.umich.edu`

Abstract

As minimum feature sizes shrink and the number of transistors integrated in a single chip grow, interconnect complexity is one of the most important issues to be solved in future VLSI chips. The use of multivalued logic is one way to effectively solve this problem because multiple-valued signals convey more information than binary signals and, thus, require a lower number of interconnecting wires to achieve similar bandwidths. This paper describes a new signed-digit adder design which uses multiple-valued logic. The circuit is composed of resonant-tunneling diodes (RTDs) and MOS transistors. The negative differential-resistance (NDR) characteristics of RTDs help to achieve very compact circuits for implementing the multiple-valued functions found in signed-digit adders. MOS transistors are useful for implementing current-mode logic, *in which addition of two or more signals is performed by simple wire interconnection. Since a redundant arithmetic is being used, the selected transfer functions allow the proposed circuit to perform addition where no ripple-carry effect is present. The design was verified using circuit simulation. To demonstrate the validity of the principles being used, a modified prototype of the circuit was built. In the prototype, a standard 2-micron CMOS process was used to fabricate the MOS-based circuitry while RTDs were connected externally. Even though no fabrication processes which integrate RTDs and MOS devices are currently available, there are efforts on the development of such technologies so that the advantages of these devices can be combined.*

1: Introduction

Devices with negative differential-resistance (NDR), such as resonant-tunneling diodes (RTDs), can be used to obtain highly compact analog and digital circuit implementations by exploiting their folded I-V characteristics [1]. Although the properties of RTDs were observed more than two decades ago [2], it was not until epitaxial growth techniques, such as molecular beam epitaxy (MBE) and metal-organic chemical vapor deposition (MOCVD), were developed that the fabrication of devices with useful NDR effects was made possible. One of the main areas of application of NDR devices is multiple-valued logic. Multiple-valued logic can help reducing interconnection complexity and chip area of integrated circuits [3], as compared with conventional binary implementations. The reason is that multivalued signals convey more information than binary signals, and progressively larger portions of chip area are used for interconnection as fabrication processes advance.

---

*This work was supported by the U.S. Army Research Office under the URI program contract DAAL 03-92-G-0109 and by ARPA under contract DAAH04-93.

0-8186-7913-1/97 $10.00 © 1997 IEEE

Signed-digit number representations were proposed in 1961 by Algirdas Avizienis [4]. This type of representations eliminate carry propagation chains by limiting the transmission of carry signals to one digit position during addition and subtraction. Signed-digit representations have an inherent redundancy that allows an addition algorithm where each digit of the resulting sum is function only of two input operand digits. The same algorithm can be used to perform addition as well as subtraction. In a redundant representation with an integer radix r each digit can assume more than r values, whereas in conventional number representations digits can only assume exactly r values. Signed-digit number representations have been used extensively. Kawahito et al. described a multiple-valued 32×32-bit multiplier [5]—one of the largest and most successful demonstrations of multiple-valued redundant-arithmetic implementations. Other important work on redundant-arithmetic implementations is presented in [6–8].

This paper presents a signed-digit adder design based on a combination of RTD and MOS transistors. The proposed design uses *current-mode* signal representation. Current-mode signal representation allows *wired addition* of signals: summing two signals simply by tying them together in a wire [9]. The adder implements radix-2, signed-digit addition. Hence, it requires a three-level logic using digits $\{-1, 0, +1\}$. The use of a three-level logic helps exploring the multiple-valued logic approach while keeping good noise margins. The proposed adder design was was verified using simulation. In order to demonstrate the validity of the proposed principle of operation, a prototype of the circuit was built. Since a standard 2-micron CMOS process was used to build the MOS portion of the circuit and RTDs were connected as discrete external elements of the test chip, it was necessary to modify the prototype to increase its flexibility and to alleviate problems caused by the incompatibility of the technologies.

Currently, it is not possible to integrate RTDs and silicon MOS transistors because RTDs are quantum devices based on compound-semiconductors such as gallium arsenide (GaAs) and indium phosphide (InP). However, various efforts aimed towards a technology which will integrate NDR and MOS elements are being conducted. One of the approaches being pursued consists of developing NDR devices made of silicon germanium (SiGe), which would make them compatible with silicon substrates [10, 11]. It is thus necessary to develop and study circuits involving RTDs and silicon MOS so that the advantages of both types of devices can be combined together.

The rest of this paper is organized as follows. Section 2 describes the principles of the redundant arithmetic being used. The design of the adder and its operation are then discussed in Section 3. Section 4 presents experimental results. Finally, conclusions are given in Section 5.

2: Signed-digit arithmetic

Signed-digit number representations were proposed by Avizienis in [4]. This type of number systems eliminate carry propagation chains by limiting the transmission of carry signals to one position during addition and subtraction. Redundancy inherent to signed-digit numbers allows an addition algorithm where each digit of the resulting sum is a function of only two operand digits. Such algorithm is useful in performing addition and subtraction operations.

Signed-digit number representations were conceived with the purpose of eliminating the propagation of the carry signal along a digital adder. In this way, the delay of the addition operation is made independent of the size of the adder. Avizienis named this type of addition *totally-parallel.* The algebraic value of a signed-digit number is given by

$$Z = \sum_{i=-n}^{m} z_i r^{-i}$$

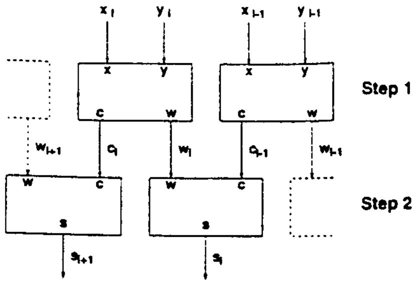


Figure 1. Block digram of the totally-parallel addition approach in signed-digit number representation.

where r is a positive integer called the *radix*. The values of the radix, r, and the number digits, z_i, should satisfy the condition of a unique representation for the algebraic value $Z = 0$; that is, the algebraic value Z is zero if, and only if, all digits of its signed-digit representation have the value $z_i = 0$. Similarly, the sign of the algebraic value Z is determined by the sign of the most significant nonzero digit. Also, the signed-digit representation of $-Z$ is obtained by changing the sign of every nonzero z_i digit.

The addition of two digits x_i and y_i is totally-parallel if two conditions are satisfied. First, the sum digit, s_i, should be function only of the operand digits, x_i and y_i, and the carry digit, c_{i-1}, from the previous digit position (see Fig. 1). Second, the carry digit to the next position, c_i, should be function only of the operand digits, x_i and y_i. Totally-parallel subtraction, $x_i - y_i$, is realized as the totally-parallel addition of x_i and the additive inverse of y_i, that is, $x_i - y_i = x_i + (-y_i)$. Fig. 1 depicts the totally-parallel addition approach in signed-digit representation.

Totally-parallel addition of two digits is performed in two steps. In the first step, a transfer digit output, c_i, and an interim sum output, w_i, are generated such that

$$x_i + y_i = rc_i + w_i. \tag{1}$$

In the second step, the final sum digit, s_i, is obtained as

$$s_i = w_i + c_{i-1}. \tag{2}$$

The required and allowed values of digits can be derived from the definition of totally-parallel addition and subtraction and from the addition algorithm described by Equations (1) and (2). For a complete analysis see [4].

The rules for obtaining w_i, c_i, and s_i can be determined, given the set of allowed values for w_i in the form of the sequence $w_{min}, \ldots, -1, 0, 1, \ldots, w_{max}$, as follows:

$$w_i = (x_i + y_i) - rc_i$$

where

$$c_i = \begin{cases} 0 & \text{if } w_{min} \leq x_i + y_i \leq w_{max} \\ 1 & \text{if } x_i + y_i > w_{max} \\ -1 & \text{if } x_i + y_i < w_{min} \end{cases}$$

and

$$s_i = w_i + c_{i-1}.$$

Signed-digit number systems have been adopted by many researchers and designers in the development of high-performance arithmetic circuits and units. One of the most successful demonstrations of multiple-valued logic is the current-mode multiplier chip developed by Kawahito et al. [5]. Kameyama presented the implementation of a multiple-valued, signed-digit arithmetic system realized in a modular fashion [8].

3: Signed-digit full adder circuit design

This section describes the operation of the proposed signed-digit full adder (SDFA) cell circuit. First, the required transfer characteristics for a three-valued arithmetic (radix-2) will be described. The circuit implementation of the identified transfer functions will then be presented. It was found that using a combination of RTD and CMOS devices offers important benefits in terms of circuit compactness in two ways. First, CMOS devices are very useful in a current-mode logic with wired-addition of digits, which saves circuit area. Second, the folded I-V characteristic of the RTD contributes with high functionality for multiple-valued logic, making the resulting design even more compact.

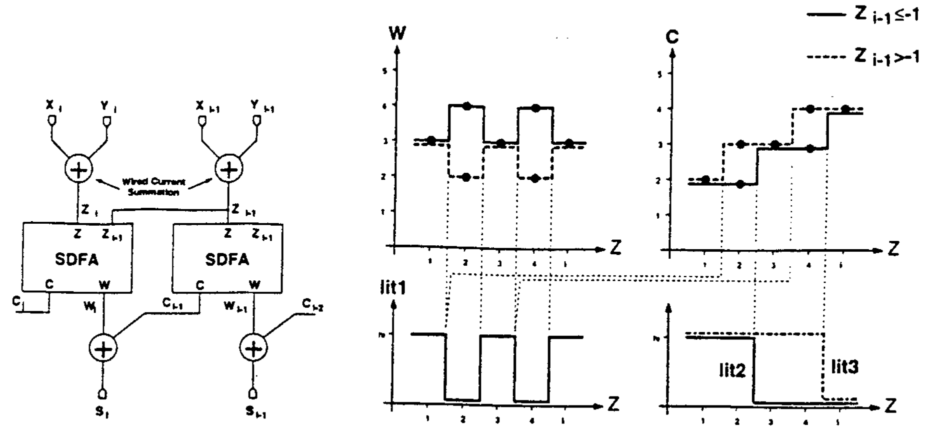
3.1: Transfer characteristics

Fig. 2(a) depicts a block diagram of the signed-digit addition approach proposed in this work. Lines x_i, y_i, c_i, w_i, and s_i are three-valued, current-mode signals. Addition of x_i and y_i is performed by simple wired-summation of currents. The function of the SDFA block is to convert the summation input signal, z, to a two-digit representation of the sum given by digits c and w; that is, $rc + w = z$ where $r = 2$. The final sum output, s_i, is obtained by current-addition of the interim sum output, w_i, and the incoming carry signal, c_{i-1}. This addition scheme clearly corresponds to the formulations stated in (1) and (2).

Note that the scheme presented in Fig. 2(a) is not identical to the one displayed in Fig. 1. The main difference is that the SDFA block requires as inputs both its corresponding input signal, z, and the input to the next less-significant adder slice, z_{i-1}. This difference in the approach is due to the use of a radix $r = 2$. According to the signed-digit arithmetic rules, the radix value should meet the condition $r > 2$. In this case, however, a *modified signed-digit* arithmetic [4] will be implemented.

The transfer functions of the SDFA block are defined so that w and c always represent the arithmetic value of $x + y$ as established in (1) and (2). Fig. 2(b) shows the transfer functions for the interim sum, w, and the carry, c, signals in the SDFA cell. All the digits in the graph are positive because the circuit will use only positive currents. In this case, the signed-digit 0 is represented by a current level "3", digit -2 is represented by current "1", and so on. There are two pairs of transfer functions, and the working pair is selected by the value of z_{i-1}. This input signal is used to determine if $c_{i-1} \neq -1$, which indicates when the SDFA cell is allowed to generate an output $w = -1$ without causing invalid s current levels to be produced. If the input z_{i-1} to the previous digit was not considered, then it would be possible to generate $w = -1$ or $w = 1$ when $c_{i-1} = -1$ or $c_{i-1} = 1$, respectively. In these cases, the final sum result would be $s = -2$ or $s = 2$, which are invalid outputs for the selected radix.

Careful observation of the SDFA transfer characteristics shows that three literal signals can be used to describe the multivalued functions. As seen in Fig. 2(b), literals *lit*1, *lit*2, and *lit*3 contain



(a)

(b)

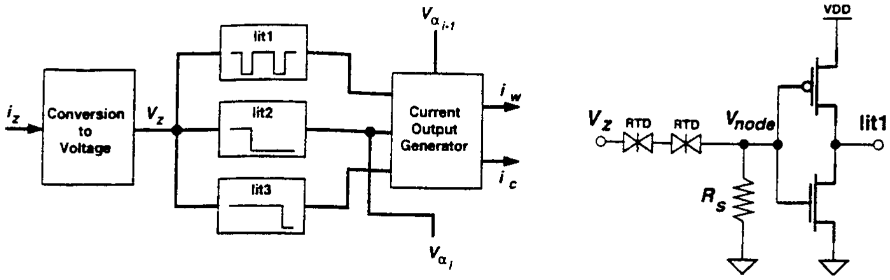
Figure 2. Totally-parallel addition approach implemented. (a) Block diagram. (b) Transfer function of SDFA block.

all the switching information required to define output functions *w* and *c*. This observation forms the basis of the proposed SDFA cell design, whose block diagram is depicted in Fig. 3(a). The input to the system is the wired addition current signal i_z. The three required literal signals are then generated in different blocks of the circuit. The literal signals are used to control switched current sources, which in turn synthesize the SDFA transfer functions in the *current output generator* block. Note that the output block uses input signal $V_{\alpha_{i-1}}$ to determine which of the two sets of transfer characteristics should be used. $V_{\alpha_{i-1}}$ indicates if $z_{i-1} < -1$ and, luckily, its behavior is identical to literal *lit*2.

3.2: Literal circuit implementation

Let's begin the description of the SDFA circuit by discussing the implementation of signal *lit*1 using RTDs. Fig. 3(b) shows the basic circuit used for generating the literal signals. As seen in the figure, the circuit consists of a series connection of the resistor R_s and the RTD, where the RTD performs as the load. The input to the circuit is the voltage V_z, which represents the wired summation $z = x + y$. In the case of *lit*1, two RTDs are used in series to obtain an equivalent two-peak characteristic. The CMOS inverter is used to sense the voltage at the node connecting the RTD and the resistor (V_{node}). The proposed literal circuit is similar to the one presented in [12]; however, the proposed design uses a complete CMOS inverter instead of a passive-load inverter.

The behavior of the voltage V_{node} is described using the load-line method (depicted in Figs. 4(a) and 4(b)). The I-V characteristics of the RTD and the resistor are plotted as the current I that flows through them with respect to the voltage V_{node}. The value of the current I is found at the point of intersection of the I-V curves. The operating point of the circuit changes as the input voltage V_z is increased because the I-V curve of the RTD is moved to the right as V_z increases. There are points where the current I suddenly decreases with increasing V_z due to the folded I-V characteristics of



(a) (b)

Figure 3. Proposed SDFA cell implementation. (a) Block diagram. (b) Compact circuit for generating *lit*1 **using RTD and CMOS devices.**

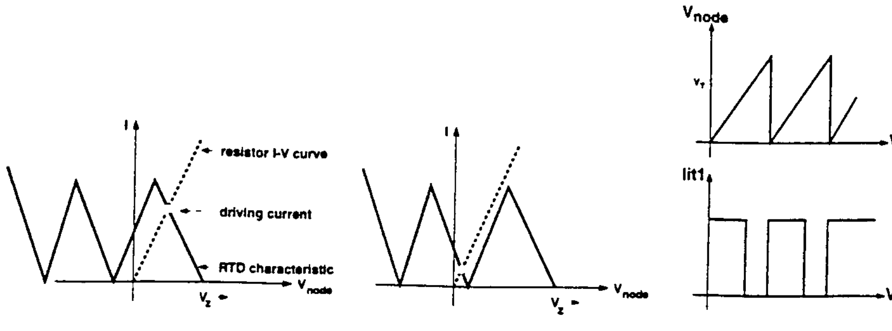
RTDs. Fig. 4(b), shows a low-current operating point occurring after a transition from a high-current point (shown in Fig. 4(a)). Since the voltage V_{node} is proportional to the current I, the transitions in I are directly reflected in V_{node}. Fig. 4(c) displays the ideal behavior of V_{node} obtained using the load-line method.

The CMOS inverter is used to sense V_{node} and then generate the transfer characteristic of *lit*1, using the threshold of the gate (see Fig. 4(c)). The CMOS inverter can be thought of as a buffer of the RTD-resistor voltage characteristic. The CMOS gate offers a very important advantage to the literal circuit: a high input resistance that prevents the alteration of the RTD characteristics caused by draining DC currents from the node V_{node}. Fig. 4(c) depicts the formation of the literal signal *lit*1 by sensing V_{node}. Observe that the values of V_z where the transition points of the literal *lit*1 take place depend on the value of the threshold voltage of the inverter (V_T).

As seen in Fig. 2(b), literals *lit*2 and *lit*3 are simpler than literal *lit*1. Consequently, implementing *lit*2 and *lit*3 is easier than implementing *lit*1 and no multipeak RTD characteristic is required. Functions *lit*2 and *lit*3 correspond to threshold detectors whose threshold levels should match the switching levels of the carry output signal, c, for $z_{i-1} \leq -1$. Fig. 6 shows the implementation of literals *lit*2 and *lit*3 by means of two resistors connected in series. The ratio between the values of the two resistors for each literal determines the corresponding switching threshold levels.

3.3: Output current circuits

The generation of the SDFA transfer functions by means of the literal signals *lit*1, *lit*2, *lit*3 is now discussed. The behavior of the literals was conceived with the purpose of achieving a simple implementation of the current output block. The approach consists of using the four controlling signals (three literals and $V_{\alpha_{i-1}}$) to activate switched current sources. These current sources are realized with MOS devices. Fig. 5 shows the current output generator circuits. Implementing output w requires only three MOS transistors and two control input signals (see Fig. 5(a)). The interim sum circuit owes its simplicity to the similarity in the forms of function w and literal *lit*1. Transistor *m*1 injects a current equivalent to one logic level when *lit*1 is high. These states correspond to $z = 1,3,5$ in the transfer function shown in Fig. 2(b). The operating points for $z = 2,4$ are handled by transistors *m*2 and *m*3, which inject a current equivalent to two logic levels or no current at all,



(a) (b) (c)

Figure 4. (a), (b) Application of the load-line method in describing the behavior of the voltage V_{node}. (c) Ideal representation of the use of the CMOS inverter in the formation of $lit1$.

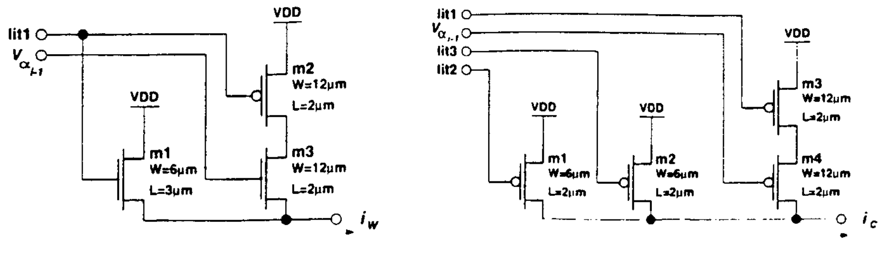
depending upon the value of $V_{\alpha_{i-1}}$.

The circuit for implementing function c is shown in Fig. 5(b). In this circuit, transistors $m1$ and $m2$ produce the basic stairway form of the carry function for the case $z_{i-1} \leq -1$ (as depicted in Fig. 2(b)). Transistors $m3$ and $m4$ are used to generate the correct carry output current, when $z_{i-1} > -1$, by injecting one logic level of current at the operating points where the two cases of the output function c are different ($z = 2, 4$). Note that $lit1$ controls these operating points because they coincide with the zero-level points of this literal signal. It is important to note that the current output functions generated by the described circuits are not identical to the ideal transfer functions shown in Fig. 2(b). The difference lies in the current output levels. While the ideal transfer function sweeps current levels 2, 3 and 4, the given output circuits generate signals that sweep levels 0, 1, and 2. This represents no problem at all. Since only positive current signals are used, it is always necessary to perform a shift correction after every wired-addition, and the amount of this correction is smaller if the lowest current levels possible are used. Fig. 6 shows a diagram of the complete SDFA circuit.

An important difference between the interim sum and carry circuits lies in the types of MOS devices they use. While the carry circuit uses only one type of transistors, the interim sum uses both nMOS and pMOS devices. Using only one type of transistors gives the carry circuit two advantages. First, the switching times of the devices are more uniform. And second, there is a better control of the current output levels because there are no different transconductance values for different types of devices. If a w current generator using only pMOS devices was required, it would be necessary to insert two inverters in the SDFA circuit to generate the complements of literal signals $lit1$ and $lit2$.

3.4: Analysis of literal circuit

The following paragraphs describe how the optimum value of the resistor R_s was determined, given its impact on performance figures and the characteristics of the RTD. The analysis assumes a single-peak RTD characteristic like the one shown in Fig. 7. Observing Fig. 4(b) one can see that,


lit1
Vα i-1
VDD
m1
W=6μm
L=3μm
m2
W=12μm
L=2μm
m3
W=12μm
L=2μm
i_w
lit1
Vα i-1
lit3
lit2
VDD
m1
W=6μm
L=2μm
m2
W=6μm
L=2μm
m3
W=12μm
L=2μm
m4
W=12μm
L=2μm
i_c

(a) (b)

Figure 5. Schematic diagram of the circuits for generating the (a) interim sum and the (b) carry output functions.

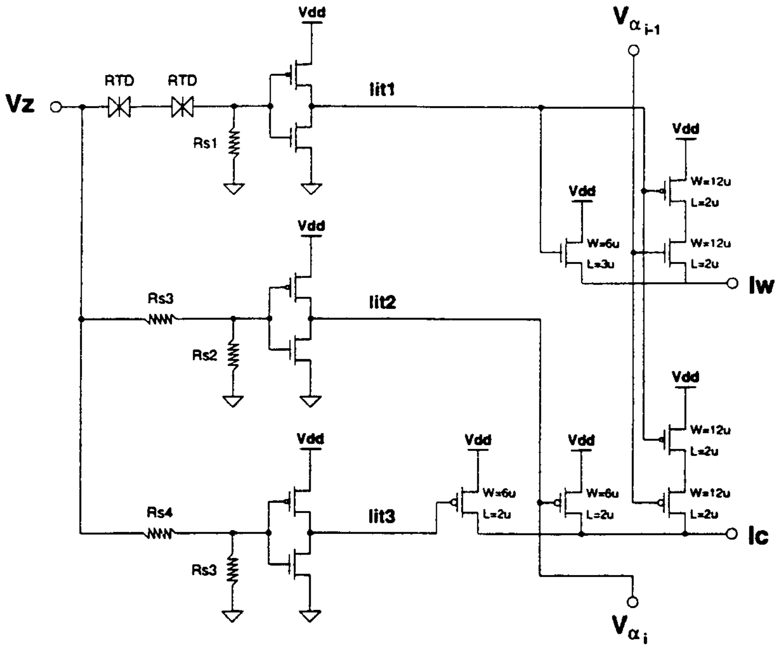

Vdd
RTD
RTD
Vz
lit1
Rs1
Vα i-1
W=12u
L=2u
W=6u
L=3u
Iw
lit2
Rs3
Rs2
Rs4
lit3
Rs3
W=6u
L=2u
Ic
Vα i

Figure 6. Circuit diagram of the complete SDFA cell.

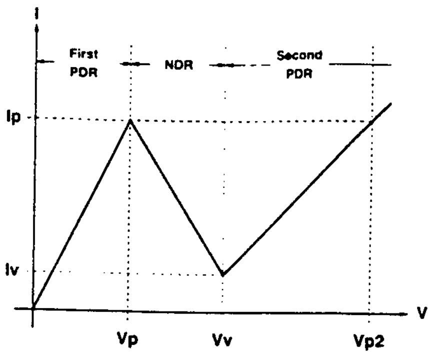


Figure 7. Current-voltage characteristic used for a single-peak RTD.

in order to have a sharp transition in V_{node}, $R_s < |R_n|$ where R_n is the negative differential-resistance of the RTD. From the characteristics, R_n can be written as

$$R_n = \frac{V_v - V_p}{I_v - I_p}$$

where V_p and V_v are the peak and valley voltages, respectively, and I_p and I_v are the corresponding peak and valley currents. In the proposed resistor-RTD topology, the maximum output voltage is given by the product of the peak current of the RTD and the value of the resistor; that is, $V_{node(max)} = I_p R_s$. On the other hand, we just saw that

$$R_s < \left| \frac{V_v - V_p}{I_v - I_p} \right| . \tag{3}$$

Therefore,

$$V_{node(max)} = I_p \frac{V_v - V_p}{I_p - I_v} .$$

Assuming a high peak-to-valley ratio ($I_p \gg I_v$), $V_{node(max)}$ can be expressed as

$$V_{node(max)} \approx V_v - V_p . \tag{4}$$

A high value of $V_{node(max)}$ is needed to switch the CMOS inverter because V_{node} has to be larger than the threshold of the inverter (V_T). Therefore, from (4), $V_v - V_p$ should be high. At the same time, from (3), $(V_v - V_p) \propto R_s$. Hence, R_s will have to be large too, which helps to improve input resistance. However, large $V_v - V_p$ is a limiting factor for RTDs and input voltage ranges. In this particular design, the RTD characteristics were predetermined by the process and the problem consisted in selecting R_s. Table 1 describes the characteristic parameters of the available RTD. In the table, *Second Voltage* is defined as the voltage $V_{p2} > V_v$ at which the current through the RTD increases to exactly I_p. C_{be1} through C_{be3} are intrinsic parasitic capacitances across the RTD.

Dynamic hysteresis is a very important factor in the selection of R_s. Dynamic hysteresis is the shifting of the RTD I-V characteristic for a dynamic input signal, due to the displacement current flowing through the parasitic capacitor C_{be} [13]. The amount of shifting (hysteresis) is larger in the NDR region than in the positive differential-resistance (PDR) region because the current flowing

Table 1. Available RTD Characteristics

Parameter	Symbol	Value
Peak Voltage	V_p	0.3 V
Peak Current	I_p	7.0 mA
Valley Voltage	V_v	0.6 V
Valley Current	I_v	0.7 mA
Second Voltage	V_{p2}	1.0 V
Voltage Difference	$V_{pp} = V_s - V_p$	0.7 V
First Positive Resistance	$R_{p1} = \frac{V_p}{I_p}$	42.9 Ω
Negative Differential Resistance	$R_n = \frac{V_v - V_p}{I_v - I_p}$	-47.6 Ω
Second Positive Resistance	$R_{p2} = \frac{V_{p2} - V_v}{I_p - I_v}$	63.5 Ω
Capacitance region 1	C_{be1}	3 fF
Capacitance region 2	C_{be2}	3 fF
Capacitance region 3	C_{be3}	2 fF

through the RTD decreases, while the displacement current tends to increase due to a positive slew rate input. The effect of dynamic hysteresis can be expressed as a delay between the expected DC I-V characteristic and the transient response for a given positive slew rate input. The amount of this delay and its relation to the value of R_s can be calculated by performing the corresponding transient analysis of the circuit. Fig. 8 depicts the equivalent circuits for each of the operating regions of the RTD. Note that the piecewise linear model of the RTD was used. In the diagrams, C_g is the load capacitance presented by the CMOS inverter of the literal circuit. With the exception of R_n, for which the negative sign is written explicitly in the circuit schematic, the circuit elements are assigned according to the RTD parameters presented in Table 1.

In order to find the delay at V_{node}, it is necessary to obtain the time $t_{dynamic}$ when $I_{R_s} = (I_p - I_v)/2$ in the NDR region of operation (Fig. 8(c)). This operating point is selected because it gives the worst case delay. The delay time t_{delay} is obtained by comparing the transient time $t_{dynamic}$ with the static time t_{static}. That is, $t_{delay} = t_{dynamic} - t_{static}$ where t_{static} is the delay for the current to reach the same operating point in the static equivalent of the circuit (without parasitic capacitances). The transient response analysis of the circuit is done in two parts. The first part is for the first PDR region of operation of the RTD, which is modeled by the circuit shown in Fig. 8(b). Analyzing the first region of operation allows the calculation of the transient time when the current through the RTD reaches I_p. This information is used to obtain the initial conditions required in the second part of the analysis. The second part is done for the NDR region of operation of the RTD, whose equivalent circuit is shown in Fig. 8(c). In this part the delay time $t_{dynamic}$, up to the operating point of interest ($I_{R_s} = (I_p + I_v)/2$), is obtained.

The first part of the analysis (PDR region) is done as follows. Refer to Fig. 8(b) for this discussion. The transient response is analyzed by solving the node equations for V_{node} in time. The flow of currents at V_{node} is written as

$$\frac{V_z(t) - V_{node}(t)}{R_p} + C_{be1}\frac{d}{dt}(V_z(t) - V_{node}(t)) = \frac{V_{node}(t)}{R_s} + C_g\frac{d}{dt}(V_{node}(t)). \qquad (5)$$

As mentioned previously, the input signal V_z is assumed to be a ramp with slew rate S_R, that is,

$$V_z(t) = S_R t. \qquad (6)$$

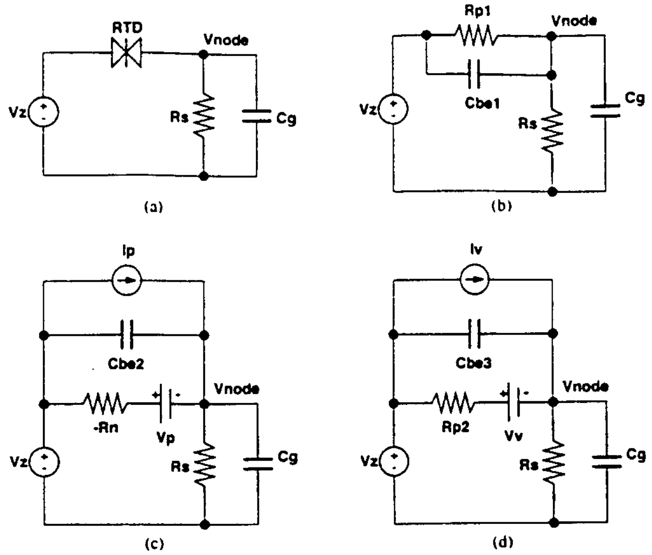


Figure 8. Equivalent circuits used in the resistor-RTD circuit transient analysis. (a) Original circuit. (b) Equivalent circuit in the first PDR region. (c) Equivalent circuit in the NDR region. (d) Equivalent circuit for the second PDR region.

Solving the linear differential equation (5) with (6) and the initial condition $V_{node}(0) = 0$, yields to the following expression for V_{node}:

$$V_{node}(t) = \frac{R_s}{R_p + R_s} V_z(t) + \frac{S_R R_p R_s}{R_p + R_s}\left[C_p - \frac{R_s(C_g + C_{be1})}{R_p + R_s}\right]\left(1 - e^{-\alpha_1 t}\right) \tag{7}$$

where

$$\alpha_1 = \frac{R_p + R_s}{R_p Rs(C_g + C_{be1})}.$$

To find the initial conditions for the second part of the analysis, expression (7) is used to solve

$$V_z(t_a) - V_{node}(t_a) = V_p$$

where t_a is the time at which the RTD enters the NDR region of operation. Using t_a the final value, V_{za}, of the input signal when the operation of the circuit leaves the PDR region ($V_{za} = S_R t_a$) is determined. V_{za} is the initial value for the input variable V_z in the second part of the analysis.

The second part of the analysis is performed in a similar fashion. In order to make things more simple, a new time reference is used. Considering this assumption and the previous result, the input variable is expressed as

$$V_z(t) = S_R t + V_{za}. \tag{8}$$

The schematic diagram in Fig. 8(c) describes the equivalent circuit for the NDR region of operation. A negative sign is written for the negative differential-resistor, and the analysis is performed accordingly. This is done with the purpose of making the negative resistor more explicit. R_n in this case is the absolute value of the negative differential-resistance presented in Table 1. It is now easy to obtain the differential equation describing the operation of the circuit at V_{node}:

$$I_p + C_{be2}\frac{d}{dt}(V_z(t) - V_{node}(t)) + \frac{V_z(t) - V_{node}(t) - V_p}{-R_n} = \frac{V_{node}(t)}{R_s} + C_g\frac{d}{dt}V_{node}(t). \qquad (9)$$

This equation is solved using the initial condition $V_{node}(0) = V_{za} - V_p$, yielding to

$$\begin{aligned} V_{node}(t) = {} & \frac{R_s}{R_s - R_n} + \\ & \frac{R_s}{R_n - R_s}\left[V_p + I_pR_n + S_RR_nC_{be2} + \frac{S_RR_nR_s(C_g + C_{be2})}{R_n - R_s}\right](1 - e^{-\alpha_2 t}) - \\ & \left(\frac{R_n}{R_s - Rn}V_{za} + V_p\right)e^{-\alpha_2 t} \qquad (10) \end{aligned}$$

where

$$\alpha_2 = \frac{R_n - Rs}{R_nR_s(C_g + C_{be2})}.$$

Expression (10) is used to calculate the time t_b when the current through the RTD/resistor reaches the operating point of interest ($I_{R_s} = (I_p + I_v)/2$). This time is obtained by solving

$$\frac{V_{node}(t_b)}{R_s} = \frac{I_p + I_v}{2}.$$

The dynamic time is obtained using t_a and t_b: $t_{dynamic} = t_a + t_b$. To calculate t_{static}, a similar type of analysis in two steps is done. The only difference is that the calculation is for the DC transfer characteristics of the RTD, hence the parasitic capacitors are eliminated. The delay time is calculated as the difference between dynamic and static times; that is $t_{delay} = t_{dynamic} - t_{static}$. Analyzing the dynamic hysteresis is useful for finding the best values for the resistor R_s. Being too difficult to obtain an analytical expression for t_{delay}, a computer was used to calculate t_{delay} for several values of R_s. The obtained values were then used to generate a graph of the variation of the delay time, t_{delay}, with respect to R_s. Fig. 9 shows the graph of t_{delay} obtained for the given RTD characteristics using the method described. This experiment showed that t_b is always very small because the transient of V_{node} in the NDR region is very sharp. Therefore, the delay time is predominantly determined by the time t_a taken to reach $I_{R_s} = I_p$.

The criteria for selecting R_s is a tradeoff between the delay and the input resistance of the literal circuit. While it is necessary to minimize the delay of the configuration, it is also necessary to maximize the input resistance. The highest value of R_s before the sharp increase in delay takes place is then selected (Fig. 9). Using this criteria and the graph, $R_s \leq 35$ ohms is considered a reasonable value. Note that the input resistance of the literal circuit is given by the addition of R_s and the positive resistance of the RTD, and not only by R_s.

4: Results

This section presents results obtained from the simulation experiments which were done to verify the proposed SDFA design. It also describes the main issues that arose when implementing the circuit prototype as well as the measured results obtained from testing the circuit.

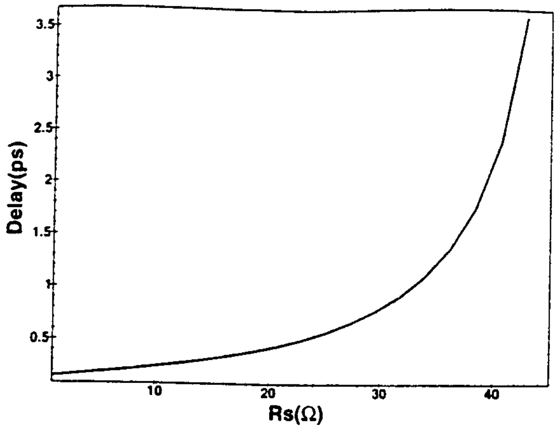
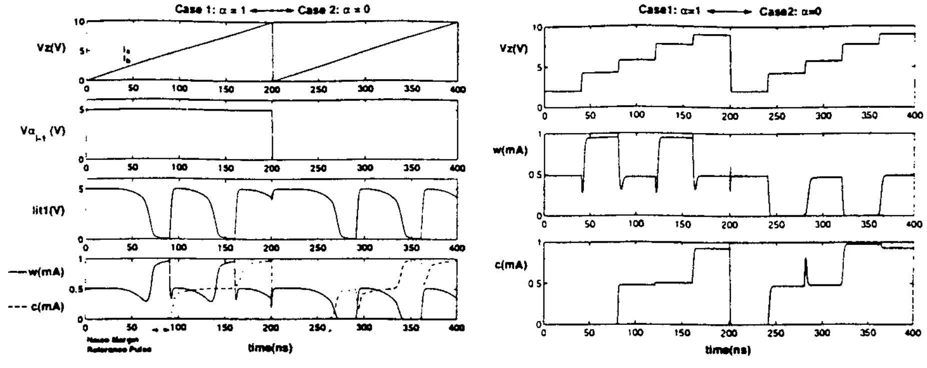


Figure 9. Plot of t_{delay} with respect to R_s for the RTD-resistor node in the literal circuit.

4.1: Simulation

The proposed SDFA circuit was verified using NDR-SPICE [14–16]. This circuit simulation tool includes SPICE models for RTDs and other NDR devices. The simulator also makes use of special convergence routines that eliminate false oscillations and other convergence problems that arise when simulating NDR-based circuits with conventional circuit simulators. Fig. 10(a) shows a transient analysis output trace obtained from a simulation of the SDFA circuit. This simulation experiment includes both cases for input $V_{\alpha_{i-1}}$: $z_{i-1} \leq -1$ and $z_{i-1} > -1$. Another important characteristic of this experiment is that the input signal is a ramp. This type of input is useful for studying the transfer characteristics of the circuit, including noise margins and identification of logic levels. Performance figures such as power consumption and delay are estimated by means of a second simulation experiment. In the second experiment, the input signal has fast transients in order to avoid contribution of the input signal rise and fall times to the delay measured at the output. In the second experiment, the input signal V_z is stepped through its different logic levels, which were determined in the first experiment. The simulation traces obtained in the second experiment are shown in Fig. 10(b).

Table 2 summarizes the results of the measurements for both simulation experiments. The noise margin was measured in the first experiment with respect to the least wide pulse in output signal w (as shown in Fig. 10(a)). The levels i_a and i_b of input signal V_z at 50% of each transition of the selected pulse of w are measured. The noise margin is obtained by assuming that the operating point is at the middle of the selected output pulse. Hence, the noise margin is half of the difference between the measured current values of signal V_Z; that is, $NM = |i_a - i_b|/2$. The value of power dissipation given in Table 2 was obtained in the second simulation experiment (Fig. 10(b)). The frequency of operation affects the value of the measured power dissipation. Finally, the delay of the circuit was measured as the time elapsed between 50% of the transition in the input signal, V_z, and 50% of the resulting transition in the output signal (in the second experiment). Both rising and falling delays were measured, but only the worst case (rising) result is given. A good characteristic of the proposed approach is that it will be able to take advantage of the progress in CMOS technology. For instance, the delay simulated values should be reduced as the circuit is implemented with



(a)

(b)

Figure 10. SDFA circuit simulation. (a) Output traces for the first experiment. (b) Output traces for the second experiment.

Table 2. SDFA Circuit Simulation Results

Parameter	Value
Noise Margin	0.15 mA
Power Dissipation	2.3 mW
Delay (sum)	3.5 ns
Delay (carry)	2.5 ns

more advanced CMOS processes.

4.2: Comparison with CMOS

The proposed signed-digit adder circuit achieves a compact implementation using 13 MOS transistors, two RTDs, and five resistors (see Fig. 6). In these terms, the proposed design is superior to the static CMOS implementation of the equivalent functional unit—the redundant binary adder cell presented in [6], which requires 44 MOS transistors. The multivalued RTD-CMOS design, however, has the disadvantage of using resistors. Even though modern technologies allow the fabrication of resistors using very small circuit areas, achieving accuracy in their values is still difficult. Eliminating this type of devices from the designs is therefore a constant objective of this research.

The main goal of this work is to demonstrate the functionality of a compact circuit for the signed-digit adder. Given the nature of the prototype—with RTDs connected externally as discrete devices—it is not possible to obtain good experimental measurements of performance figures such as power consumption and propagation delay. Hence, it is not yet possible to provide accurate comparisons of performance of the proposed approach against that of conventional technologies such as CMOS. It is important to note that the signed-digit addition approach is intrinsically fast because it avoids propagation of carry signals. Consequently, the improvement in speed will be higher when many signed-digit adder cells are put together to form a parallel addition unit. In such

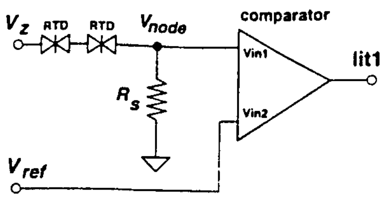


Figure 11. Modified circuit for generating *lit*1**.**

parallel adder, the computation time will be constant, while in conventional approaches the delay is a function of the word-length of the operands. To make a good comparison of performance, it is thus necessary to compare wide parallel adders instead of single adder cells.

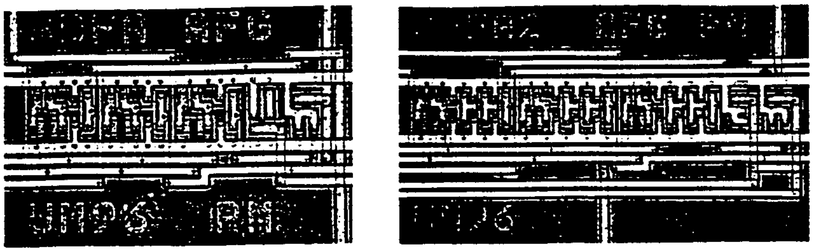
Table 2 presents the values for delay and power dissipation obtained from the circuit simulation experiments. Being a current-mode circuit, the proposed approach is obviously not competitive with CMOS in terms of power dissipation. The measured worst-case delays are similar to the types of delays that could be obtained in the CMOS redundant binary adder. However, the simulation experiments did not take the effects of interconnecting wires into account. It is well known that interconnect delays will significantly determine the performance of submicron VLSI and ULSI circuits [17], and multivalued logic will be very useful in reducing the length of interconnecting wires [18].

4.3: Prototype implementation

It is important to implement a working prototype in order to demonstrate the principle of operation of the approach. With this purpose in mind, a small test circuit was designed using a standard 2-micron CMOS process. This being the initial effort in the construction of a prototype and functional testing the main objective, the RTD and resistor elements were connected as discrete external devices to the CMOS test chip. It was also necessary to make modifications on the prototype due to incompatibilities of the CMOS technologies and the available RTDs. Based on Equation (4) and Table 1, the highest voltage that can be obtained at V_{node}, in Fig. 3(b), is 0.3 volt. This voltage is very low compared to the approximately 2.5-volts threshold of an inverter in the available CMOS process, where $V_{DD} = 5$ volts. In the future, RTD and CMOS characteristics will be brought closer together by CMOS scaling and the use of lower power supply voltages. However, given the physical limitations of MOS devices, it will also be necessary to modify the characteristics of RTDs.

For the SDFA prototype, it was necessary to modify the implementation as follows. The modification consisted of replacing the inverter of the literal circuit shown in Fig. 3(b) by a voltage comparator, as shown in Fig. 11. This is not the most compact option for solving the problem, but it is the most flexible. Instead of altering the threshold voltage of the inverter by using circuit techniques, such as replacing the pMOS transistor of the inverter by a resistive load, the design allows for the implementation of an adjustable threshold voltage through an external reference, V_{ref}. This approach has the disadvantage of using more transistors and a reference voltage. On the other hand, it provides the important benefit of increased flexibility. Therefore, the use of a comparator is considered advantageous and appropriate for the test chip prototype. Nevertheless, mature implementations should aim to use the simple inverter.

Fig. 12 shows microphotographs of two SDFA cells included in the chip. The first SDFA cell, shown in Fig. 12(a), is implemented using no literal complements and the sum current generator



(a) (b)

Figure 12. Microphotographs of the fabricated test chip. (a) MOS circuitry for the first version of the SDFA cell. (b) MOS circuitry for the second version of the SDFA cell.

that uses both nMOS and pMOS transistors (shown in Fig. 5(a)). The second SDFA cell, depicted in Fig. 12(b), generates literal complements and uses pMOS-only current output generators.

4.4: Experimental results

Since the main objective of the prototype to demonstrate the functionality of the SDFA cell, an experiment to show the transfer characteristics of the circuit was performed. These transfer characteristics were obtained by feeding the input of the circuit with a ramp signal of very low slew-rate. The experiment was done using the second version of the SDFA cell (Fig. 12(b)). The oscilloscope traces obtained in the experiment are shown in Fig. 13. In general, the form of the measured output functions agree with the expected circuit behavior shown in Fig. 10. Note that the experiment is divided into two parts with respect to the value of $V_{\alpha_{i-1}}$. The trace labeled "V_{node}" reflects the operation of the RTD and corresponds to the expected behavior seen in Fig. 4(c). Also, note that the behavior of literal *lit*2 is independent of the value of $V_{\alpha_{i-1}}$. The last two traces show the transfer characteristics for the output functions, *w* and *c*, of the SDFA cell. Observe how the output characteristics are selected by $V_{\alpha_{i-1}}$.

There is a difference in the values of the output current levels between the measured characteristics and their expected values. While the expected output currents for logic levels "-1", "0", and "1" are 0.0, 0.5, and 1.0 milliampere, respectively, the measured current levels are 0.0, 0.2, and 0.4 milliampere (Fig. 13). This alteration of the current levels is due, in part, to a reduction in the transconductance of the pMOS devices—from 20.563 $\mu A/V^2$ in the SPICE model used for the simulations, to 16.129 $\mu A/v^2$ measured for the actual CMOS run of the test chips. Another reason for the difference in output levels is due to the method used for measuring the output currents. In the experiment, a resistive load was connected to the outputs in order to convert the current signals into voltage signals that could be displayed by the oscilloscope.

Noise margins were measured in the same way as it was done in the simulation experiments. The worst case margin is obtained by choosing the narrowest output pulse. In this experiment, the values of the transition input currents were measured for the four pulses in the sum output function. Table 3 presents the results of the noise margin measurements. An important difference with respect to the simulation experiment is that, in the circuit prototype, the worst case pulse was the fourth pulse instead of the first one (shown in Fig. 10(a)). Note that the results are similar to those obtained

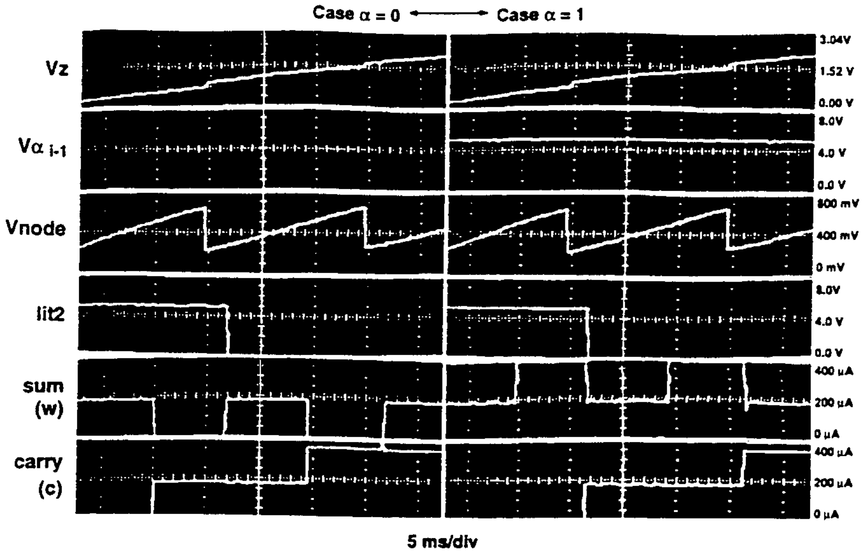


Figure 13. Oscilloscope traces showing the operation of the signed-digit adder test circuit at low frequency.

Table 3. Measured SDFA Circuit Noise Margins

Pulse	NM (mA) (α = 0)	NM (mA) (α = 1)
1st	0.2389	0.2389
2nd	0.1592	0.1592
3rd	0.2389	0.2389
4th	0.1433	0.1433

in simulation. It is important to observe that the measured noise margins were scaled down by the new current levels obtained at the outputs (by assuming matching input and output levels). Another interesting result observed in Table 3 is the constant value of noise margins in both output characteristics (irrespective of the value of $V_{\alpha_{i-1}}$). This behavior was expected because input $V_{\alpha_{i-1}}$ is connected only to the current output circuitry. Therefore, $V_{\alpha_{i-1}}$ does not affect the switching thresholds of the output function, which are defined in the input block and the literal generator circuits.

5: Conclusions

A novel multiple-valued signed-digit adder implementation is presented which combines the advantages of resonant-tunneling diodes (RTDs) and MOS transistors. RTDs provide high functionality for multiple-valued logic, and MOS transistors allow efficient arithmetic circuit designs through current-mode operation. The principle of operation of the proposed design was demonstrated by computer-based circuit simulation and by means of a modified prototype of the circuit. The results obtained indicate that RTD+CMOS, and other NDR+CMOS, technology will be very useful in the development of compact implementations of multiple-valued logic functions. The results also indicate that it is necessary to pay special attention to the compatibility between the I-V characteristics of RTDs and MOS devices in the new technologies being developed. Better matching characteristics will be obtained with scaled-down MOS processes. However, due to physical limitations on the threshold voltages of MOS transistors, it is also necessary to modify RTD characteristics so as to bring them closer to those of MOS devices.

References

[1] F. Capasso, S. Sen, F. Beltram, L. M. Lunardi, A. S. Vengurlekar, P. R. Smith, N. J. Shah, R. J. Malik, and A. Y. Cho, "Quantum functional devices: Resonant-tunneling transistors, circuits with reduced complexity, and multiple-valued logic," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2065–2082, October 1989.

[2] L. L. Chang, L. Esaki, and R. Tsu, "Resonant tunneling in semiconductor double barriers," *Applied Physics Letters*, vol. 24, pp. 593–595, 1974.

[3] K. C. Smith, "The prospects for multivalued logic: A technology and application view," *IEEE Transactions on Computers*, vol. C-30, pp. 619–634, September 1981.

[4] A. Avizienis, "Signed-digit number representations for fast parallel arithmetic," *IRE Transactions on Electronic Computers*, pp. 389–400, September 1961.

[5] S. Kawahito, M. Kameyama, T. Higuchi, and H. Yamada, "A 32×32-bit multiplier using multiple-valued MOS current-mode circuits," *IEEE Journal of Solid State Circuits*, vol. 23, pp. 124–132, February 1988.

[6] H. Makino, Y. Nakase, H. Susuki, H. Morinaka, H. Shinohara, and K. Mashiko, "An 8.8-ns 54 × 54-bit multiplier with high speed redundant binary architecture," *IEEE Journal of Solid State Circuits*, vol. 31, pp. 773–783, June 1996.

[7] M. Kameyama, T. Sekibe, and T. Higuchi, "Highly parallel residue arithmetic chip based on multiple-valued bidirectional current-mode logic," *IEEE Journal of Solid State Circuits*, vol. 24, pp. 1404–1411, October 1989.

[8] M. Kameyama, M. Nomura, and T. Higuchi, "Modular design of multiple-valued arithmetic VLSI system using signed-digit number system," in *Proceedings of the International Symposium on Multiple-Valued Logic*, pp. 355–362, IEEE, 1990.

[9] M. Kameyama and T. Higuchi, "Design of a radix-4 signed-digit arithmetic circuit for digital filtering," in *Proceedings of the International Symposium on Multiple-Valued Logic*, pp. 272–277, IEEE, 1980.

[10] U. König, M. Kuisl, J.-F. Luy, and F. Schäffler, "Si/SiGe resonant tunneling devices separated by surrounding polysilicon," *Electronics Letters*, vol. 25, pp. 1169–1171, 1989.

[11] U. König, M. Kuisl, F. Schäffler, G. Fischer, and T. Kiss, "Operating CMOS after a Si-MBE process: A precondition for future three-dimensional circuits," *IEEE Electron Device Letters*, vol. 11, pp. 218–220, May 1990.

[12] T. Hanyu, Y. Yabe, and M. Kameyama, "Multiple-valued programmable logic array based on a resonant tunneling diode model," *IEICE Transactions on Electronics*, vol. E76-C, pp. 1126–1132, July 1993.

[13] S.-J. Wei, H. C. Lin, R. C. Potter, and D. Shupe, "Dynamic hysteresis of the rtd folding circuit and its limitation on the a/d converter," *IEEE Transactions on Circuit and Systems—II: Analog and Digital Signal Processing*, vol. 39, pp. 247–251, April 1992.

[14] S. Mohan, J. P. Sun, P. Mazumder, and G. I. Haddad, "Device and circuit simulation of quantum electronic devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, pp. 653–662, June 1995.

[15] S. Mohan, P. Mazumder, and G. I. Haddad, "NDR SPICE: A circuit simulator for resonant tunneling devices," in *Proceedings of the International Compound Semiconductors Conference*, September 1994.

[16] S. Mohan, P. Mazumder, and G. I. Haddad, "A new circuit simulator for negative resistance devices," in *Proceedings of the International Electron Devices Meeting*, December 1994.

[17] M. Shoji, *High-Speed Digital Circuits*, ch. 2, pp. 73–135. Addison-Wesley, 1996.

[18] T. Hanyu, M. Kameyama, and T. Higuchi, "Prospects of multiple-valued VLSI processors," *IEICE Transactions on Electronics*, vol. E76-C, pp. 383–391, March 1993.

Design and Analysis of a Novel Quantum-MOS Sense Amplifier Circuit

Tetsuya Uemura
Fundamental Research Laboratories,
NEC Corporation
34 Miyukigaoka, Tsukuba, Ibaraki 305-8501
uemura@frl.cl.nec.co.jp

Pinaki Mazumder
Electrical Engineering and Computer Science,
The University of Michigan
Ann Arbor, MI 48109
mazum@eecs.umich.edu

Abstract

A novel quantum-MOS sense amplifier circuit consisting of resonant tunneling diodes (RTD's) as pull-up devices and NMOS transistors is discussed in this paper. Compared to the conventional sense amplifier circuits using CMOS technology, the proposed QMOS sense amplifier exhibits about 20% higher sensing speed. The cross-coupled QMOS latch, which is at the heart of the sense amplifier circuit, has metastable and unstable states which are closely related to the I-V characteristics of the RTD's. The stability analysis has been made by using phase-plot diagram and how RTD parameters relate to circuit speed and robustness of the sense amplifier has been discussed.

1 Introduction

Design of the sense amplifier is one of the most critical and difficult tasks in a MOS implementation of high-density, high-speed dynamic random access memory (DRAM) chip. In conventional DRAMs, the sense amplifier circuit usually consists of a cross-coupled complementary MOS inverter pair (latch) since it has several merits including increased chip reliability, higher speed, lower transient power dissipation, and reduced voltage bounce noise, in comparison with purely n-type FET based sense amplifiers [1]. But the main drawback of a CMOS sense amplifier is its large area owing to bulky p-type MOS transistors and this poses a severe challenge in modern high-density DRAMs with low inter-cellar pitch width (about 3 λ). Optimization of the size of PMOS devices can only be accomplished at the cost of sensing speed and noise performance. Also, larger PMOS transistors tend to increase the bit-line RC delays due to increasing bit-line lengths. It is not entirely clear whether a good compromise between disparate conflicting requirements in the design of a sense amplifier can be achieved using the conventional CMOS technology. This paper introduces a new quantum-MOS technology (QMOS) and it shows how a good sense amplifier design can be made using QMOS circuits that do away with bulky PMOS transistors. QMOS technology co-integrates quantum devices such as the resonant tunneling diodes (RTD's) and NMOS transistors. This emerging technology has a large potential and may define the future of very large scale integrated (VLSI) circuits with faster speed as well as lower dynamic power dissipation as compared to the conventional CMOS circuits [2]. Wagt, *et. al.* has proposed a revolutionizing DRAM cell design that employs a pair of RTDs to store the cell information, and thereby decreases the effect of charge leakage in the storage capacitor, which, in turn, has been shown to decrease the power consumption of a DRAM chip by over one thousand times due to the elimination of cell refreshing [3]. In this paper, a novel RTD-based sense amplifier circuit, called a quantum MOS (QMOS) sense amplifier, is proposed and its operation is analyzed and verified through SPICE simulation. The goal of this work is to demonstrate how RTDs can make a great impact in the design and growth of future DRAM chips that will consume less power and will have improved memory cycle due to the co-integration of RTDs in memory array storage cells and also in sense amplifiers.

2 Operation Principle

Figure 1(a) shows a conventional cross-coupled CMOS inverter latch. The bit lines are precharged to half-V_{DD} since the CMOS inverter exhibits a high gain in its transient region. In the *READ* operation, the charge sharing between the bit line and the accessed cell induces a small voltage difference between BIT and $\overline{BIT}$ nodes. Once the voltage difference is established, the control signals, SE and $\overline{SE}$, activate the cross-coupled latch, and the small voltage difference is amplified to push the BIT voltage to either V_{DD}

0-7695-0104-4/99 $10.00 © 1999 IEEE

[illegible] on the sign of initial voltage difference. [illegible] shows the proposed QMOS sense amplifier [illegible] CMOS inverter is replaced by the QMOS in[illegible] consists of an NMOS driver and an RTD load. As [illegible] circuit. once the voltage difference between [illegible] nodes is built, the QMOS sense amplifier is [illegible] switching transistors, $M5 - M8$. Because [illegible] current drive capability of the RTD load com[illegible] PMOS transistor, the QMOS sense amplifier is [illegible] a faster sensing time.

[illegible] Simulation Results

[illegible] 2 and 3 show a load line diagram of QMOS in[illegible] SPICE simulation result of sensing time of the [illegible] inverter, respectively. The circuit equation [illegible] given by,

$$\text{[illegible]} = I_{RTD}(V_{DD} - v_1) - I_{TR}(v_2, v_1) \quad (1)$$

$$\text{[illegible]} = I_{RTD}(V_{DD} - v_2) - I_{TR}(v_1, v_2) \quad (2)$$

[illegible] the bit-line capacitance, v_1 and v_2 are BIT [illegible] voltages. The $I_{RTD}(v)$ and $I_{TR}(v_g, v_{ds})$ indicate [illegible] voltage characteristics of the RTD and NMOS [illegible] respectively, which are given by,

$$\begin{aligned} \text{[illegible]} & \qquad (0 \le v < V_P) \\ = \text{[illegible]} - \frac{I_P - I_V}{V_V - V_P}(v - V_P) & \qquad (V_P \le v < V_V) \\ \text{[illegible]} - \frac{I_P - I_V}{V_{SEC} - V_V}(v - V_V) & \qquad (V_V \le v) \end{aligned} \quad (3)$$

[illegible]

$$\text{[illegible]} = \begin{cases} \text{[illegible]} & (v_g < V_T) \\ \text{[illegible]}v_g - V_T)v_{ds} - v_{ds}^2/2] & (v_{ds} < v_g - V_T) \\ \text{[illegible]}v_g - V_T)^2 & (v_{ds} \ge v_g - V_T) \end{cases} \quad (4)$$

[illegible] V_V, and V_{SEC} represent a peak [illegible] peak voltage, a valley current, a valley voltage [illegible] peak voltage for the RTD characteristics, re[illegible]. The [illegible] and V_T are gain and threshold voltage, [illegible] the NMOS transistor. The RTDs with [illegible] = 0.8V, $I_V = 77.4\mu$A, $V_V = 1.92$V, and [illegible] and the transistors with $\beta = 2.0 \times 10^{-4}$A/V² and V_T = [illegible] are used in the proposed design. The ini[illegible] the [illegible] voltage is 1.8 V and 1.7 V, and its [illegible] assumed to be 0.5 pF.

[illegible] in Fig. 3 indicates the sensing time of [illegible] amplifier shown for the comparison. The NMOS characteristics are assumed to be the same in the both sense amplifier circuits. The PMOS characteristics of the CMOS sense amplifier is assumed to be symmetric to the NMOS characteristics($\beta_P = \beta_N$, $V_{TP} = -V_{TN}$). The sensing time, defined as a time at which both conditions $v_1 > 0.9V_{DD}$ and $v_2 < 0.1V_{DD}$ are simultaneously satisfied, is about 20% faster than that of the CMOS. This is due to the larger current drive capability of the RTD load than the PMOS. As the current drive capability of the pull-up load is larger. the charging time of the bitline capacitance becomes shorter. This fast charging time induces the reduction of discharging time of other bitline capacitance also, because the NMOS transistor of the discharging circuit quickly turns on due to the cross coupled connection.

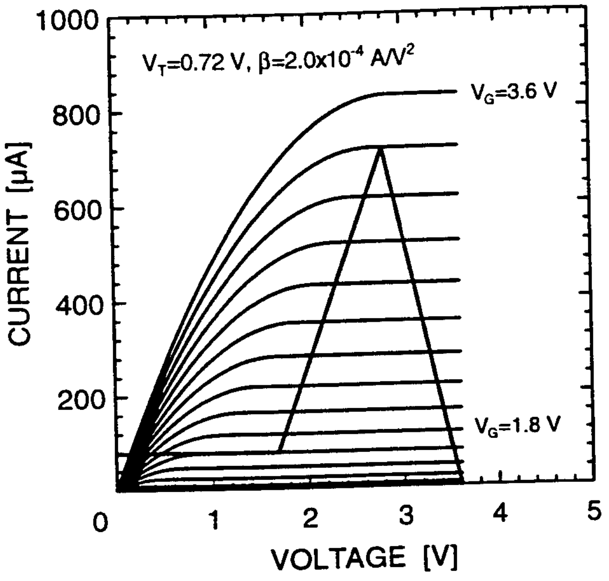


Figure 2. Load line diagram for QMOS inverter circuit

4 Stability Analysis

In this section, the stability of the solution of Eqs. 1 and 2 for a given initial condition is discussed by using phase diagram. Figure 4 shows voltage transfer characteristics (VTC) of each inverter. The solid and dashed VTC lines indicate the solution for $dv_2/dt = 0$ and $dv_1/dt = 0$, respectively. The intersection points, therefore, represent the equilibrium points of Eqs. 1 and 2. Unlike the CMOS sense amplifier, the QMOS sense amplifier circuit has three stable points (A, B, and C), one meta-stable point (D), and one unstable point (E). due to its Z-shaped VTC, as shown in Fig. 4. The dotted lines in Fig. 4(b) indicate the trajectory curves of Eqs. 1 and 2 for different initial condi-

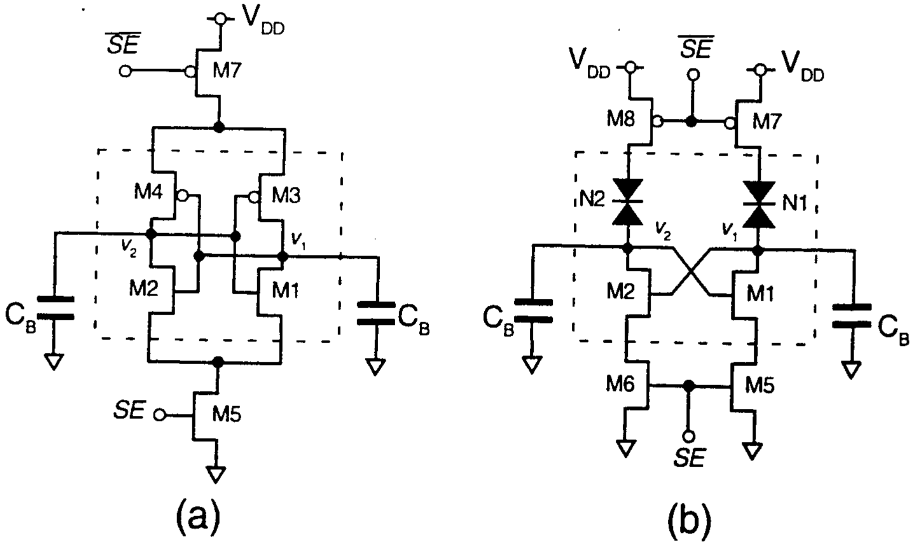


Figure 1. Sense amplifier circuit. (a) CMOS, (b) QMOS

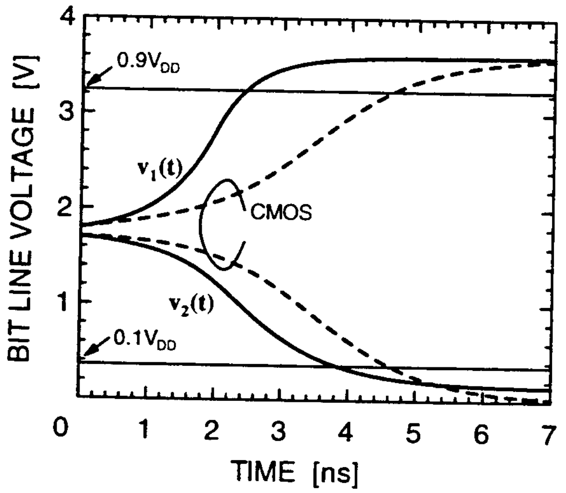


Figure 3. Simulation results for QMOS and CMOS sense amplifier circuit

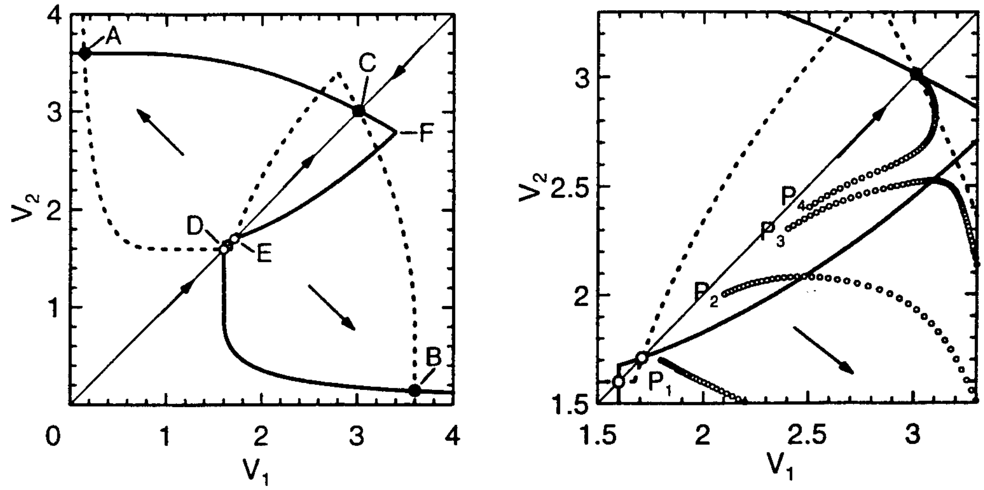
tions. From the point $P_1(1.8, 1.7)$, where $dv_1/dt > 0$ and $dv_2/dt < 0$, the solution directly goes to the stable point B (Fig. 3). From $P_2(2.0, 1.9)$ or $P_3(2.4, 2.3)$, where both dv_1/dt, and dv_2/dt, > 0, the trajectory once goes upward, and when it reaches to the solid VTC curve, it turns to the downward, resulting in going to the point B. When it starts from the point $P_4(2.5, 2.4)$, however, the solution reaches to the dashed VTC curve, resulting in convergence to the point C, which is an erroneous operation of the sense amplifier.

This erroneous behavior can be controlled by optimizing the RTD parameters. If the point F on the solid VTC curve, corresponding to the peak position of RTD characteristics, is moved closer to the line of $v_2 = v_1$, the instability is reduced. The v_1 and v_2 value at point F, which is determined by the peak current and voltage of the RTD characteristics, is given by,

$$\begin{aligned} v_1|_F &= V_T + \sqrt{\frac{2I_P}{\beta}} \\ v_2|_F &= V_{DD} - V_P \end{aligned} \qquad (5)$$

From the above equations, when I_P is decreased, the point F is moved to left. When V_P is decreased, on the other hand, the point F is moved upward. Therefore, the decrease of I_P and/or V_P can reduce the instability. Figure 5 shows an initial condition dependence of the sensing time for three different peak voltage values (V_P =0.8, 0.6, and 0.4V). The initial condition for $v_1(t)$ and $v_2(t)$ are assumed to be,

$$\begin{aligned} v_1(0) &= 1.9 + \Delta v \\ v_2(0) &= v_1(0) - 0.1 \end{aligned} \qquad (6)$$



(a)

(b)

Figure 4. Voltage transfer characteristics of QMOS inverter pair

where Δv indicates the voltage fluctuation during the bitline precharge cycle. As shown in Fig. 4 (b), when Δv equals to 0.5V at $V_P = 0.8$V, which corresponds to the point P_4 in Fig. 4 (b), the sensing time becomes infinity due to the erroneous operation. Figure 5 indicates that the reduction of V_P can increase the normal operation margin against the initial voltage fluctuation without degrading the sensing speed.

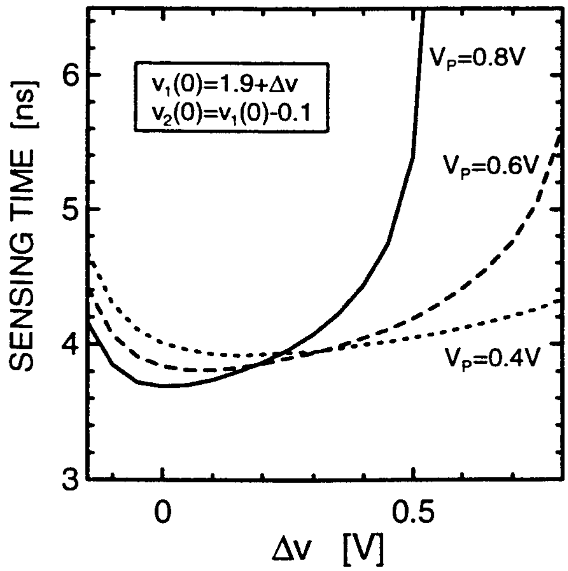


Figure 5. Initial voltage fluctuation dependence of the sensing time

5 Conclusion

Design of a novel QMOS sense amplifier circuit consisting of RTD pull-up loads and NMOS transistors is discussed in this paper and its operation is analyzed. Compared to the conventional CMOS sense amplifier, the QMOS design exhibits about 20% higher sensing speed due to the larger current drive capability of the RTD load. The instability against the initial condition has been analyzed by drawing phase plot diagram. It has been found that by reducing the peak current and/or peak voltage of the RTD circuit instability can be tackled.

References

[1] N. C. Lu, and H. H. Chao. *IEEE J. Solid-St. Circ.*, SC-19(4):451-454, August 1984.

[2] P. Mazumder, S. Kulkarni, M. Bhattacharya, J. P. Sun, and G. I. Haddad. *PROC. IEEE*, 86(4):664-685, April 1998.

[3] P. van der Wagt, A. Seabaugh, and Ed Beam III. *1996 IEDM Tech. Dig.*, IEDM-96:425-428, December 1996.

Epitaxial growth of ZnS on bare and arsenic-passivated vicinal Si(100) surfaces

Xiaochuan Zhou, Shan Jiang, and Wiley P. Kirk[a]
NanoFAB Center, Engineering-Physics Building, Texas A&M University, College Station, Texas 77843-4242

(Received 11 November 1996; accepted for publication 14 May 1997)

We report a detailed study of molecular beam epitaxial growth of ZnS films on bare and arsenic-passivated vicinal Si(100) surfaces. This study elucidates the initiation of microtwinning and stacking-fault defects on double-stepped substrate surfaces. The study also sheds light on the function of arsenic passivation in reducing crystal defects in ZnS epitaxial layers. Three substrate surfaces, Si(100) 2×1, Si(100):As 2×1, and Si(100):As 1×2, were used for the ZnS epitaxial growth studies. Adsorption experiments were performed to demonstrate the chemical passivation effect of an arsenic overlayer. Reflection high-energy electron diffraction was used to study growth modes and the epitaxial relationship of the ZnS layers to the substrates. Transmission electron microscopy was used to study the crystal-defect structures. Secondary ion mass spectroscopy was used to determine the chemical profiles of the heteroepitaxial interfaces of ZnS layers grown on arsenic-passivated surfaces. One of the main results demonstrated by this work is that thin ZnS films can be grown epitaxially with much better crystal quality on As-passivated Si surfaces than on bare Si surfaces. © *1997 American Institute of Physics.* [S0021-8979(97)02916-2]

I. INTRODUCTION

The growth of compound semiconductors on silicon substrates will be a key enabler for silicon-based electronic and optoelectronic devices.[1] While many studies have been performed using silicon as an epitaxial substrate, there have been few studies that explored the possibility of using silicon in the active layers of a heterostructural device.[2,3] One such device is the double-barrier resonant-tunneling diode (DB-RTD). In order to achieve room-temperature quantum confinement of carriers in a silicon-based DB-RTD, and minimize valley current, a wide-band-gap semiconductor is required as the barrier material. One candidate for a barrier material is zinc sulfide, which has a band gap of 3.68 eV, and a zinc-blende (cubic) crystal structure with a lattice constant of 5.420 Å.[4] By comparison, silicon has a band gap of 1.12 eV and a diamond (cubic) crystal structure with a lattice constant of 5.431 Å.[4] The band offset of a molecular beam epitaxy (MBE) grown ZnS/Si(111) heterojunction was measured by Maierhofer *et al.*,[5] using photoelectron spectroscopy. They reported a valence-band offset of $\Delta E_v = 0.7$ eV and a conduction-band offset of $\Delta E_c = 1.7$ eV. To our knowledge, a reliable experimental value for the band offset of the ZnS/Si(100) heterojunction does not exist. The closely matched lattice constants (0.2% lattice mismatch) and the large band-gap difference between the two semiconductors motivated the present study of epitaxial growth of ZnS on silicon substrates.

Although ZnS and Si are closely lattice matched, previous efforts to grow ZnS on Si substrates did not produce device quality materials.[6] The problem was partially attributed to chemical reactions during initial ZnS growth on silicon substrates.[7,8] In a photoemission surface core-level study of sulfur adsorption on Si(100), Weser *et al.*[9] found that sulfur penetrated into the crystal volume and formed disordered silicon sulfide. In a study of a similar material system involving ZnSe growth on Si(100), Bringans *et al.*[10] found the formation of an amorphous $SiSe_2$ layer at the ZnSe/Si interface. The amorphous layer disrupted the epitaxial template and caused a significant number of crystal defects to form in the deposited ZnSe films. They also showed that the deposition of an arsenic monolayer on a Si(100) surface could effectively passivate the surface and prevent a chemical reaction between selenium and silicon, thus, improving the crystalline quality of the ZnSe layers. In our earlier study of ZnS growth on Si(100), a significant improvement of ZnS crystal quality was also demonstrated by using the arsenic passivation method.[8] While the modified chemical property of the substrate surface was considered to be responsible for the improvement of ZnS and ZnSe crystal qualities, the detailed mechanism was not understood.

In this paper, we report the results of a detailed study of ZnS growth on bare and arsenic-passivated Si(100) surfaces to elucidate the mechanism of crystal-defect formation in ZnS layers grown directly on bare silicon substrates and the function of an arsenic overlayer in reducing crystal-defect formation. We used residual gas analysis to study the adsorption and desorption of sulfur and zinc on clean and arsenic covered silicon surfaces. We used reflection high-energy electron diffraction (RHEED) and secondary ion mass spectroscopy (SIMS) to examine the atomic structures of the epitaxial surfaces and the chemical profiles of the epitaxial layers. We also used transmission electron microscopy (TEM) to study the crystal structure of the epitaxial materials.

II. EXPERIMENT

Epitaxial growth was carried out in a Fisons (VG) V80S MBE system. The system was equipped with an e-beam evaporator for silicon deposition and three *K* cells: a compound ZnS source and elemental sources of arsenic (As_4) and zinc. The compound ZnS source produced stoichiometric atomic/molecular beam fluxes of Zn and S_2. The MBE system was also equipped with a 15 keV RHEED system and a Dycor M200 residual gas analyzer (RGA) system to do surface structural studies and gas-phase composition analysis, respectively. The RGA was positioned in the line of sight of the substrate and had an orifice to collimate the gas molecules that were backscattered or desorbed only from the substrate surface.

---

[a] Electronic mail: Kirk@nanofab.tamu.edu

Vicinal Si(100) substrates with a 4° offcut towards [011] were used to form a double-stepped surface to suppress antiphase disorder. The silicon substrates were chemically cleaned using a peroxide cleaning procedure, which left surfaces covered with a thin layer of oxide formed in the $HCl:H_2O_2:H_2O=3:1:1$ solution.[11] The substrates were annealed in the MBE preparation chamber at 870 °C for 20 min and then transferred into the growth chamber. Streaky RHEED patterns from the 2×1 reconstructed Si(100) surfaces were observed. Silicon layers of 100 to 1000 Å were grown on the substrates at a temperature of 600 °C to achieve a smooth starting surface.

The ZnS growth was usually done in two steps. The first step was performed at a low substrate temperature between 50 and 220 °C and a low $Zn+S_2$ beam-equivalent pressure of 1×10^{-7} mbar for 5 min. These lower growth temperatures resulted in smoother surface morphologies. In the second step, the growth was done at a substrate temperature of 300 °C and a $Zn+S_2$ beam flux pressure of 1×10^{-6} mbar.

Gas-phase analysis was performed by monitoring three mass-to-charge ratios, $m/e=60$, 64, and 66, with the RGA. Individual partial pressures of the SiS, S_2, and Zn were calculated, based on the natural abundance of the silicon, sulfur, and zinc isotopes. An ion gauge, which was located in front of the substrate surface, was used to calibrate the RGA pressure measurement.

Cross-sectional TEM samples were prepared by using standard mechanical cutting, lapping, and ion-milling procedures. All TEM micrographs were made using a JEOL 2010 microscope at 200 kV.

III. RESULTS

A. Structure of substrate surfaces

In order to grow antiphase-free ZnS layers on silicon surfaces, we used vicinal Si(100) substrates offcut 4° towards [011], so that, with proper substrate treatment, double-stepped surfaces could be formed.[12] After growing a thin buffer layer of silicon, we annealed the substrates at 780 °C for 20 min and then gradually decreased the substrate temperature below 500 °C. Figures 1(a) and 1(b) show the RHEED patterns of such an annealed vicinal Si(100) surface in the [011] and $[0\bar{1}1]$ azimuthal directions, respectively. A 2×1 reconstructed single-domain surface was observed. This RHEED pattern corresponds to a double-stepped surface with dimer bonds parallel to the step edges.[13] This was the starting surface we used for epitaxial growth of ZnS films on a bare Si(100) surface.

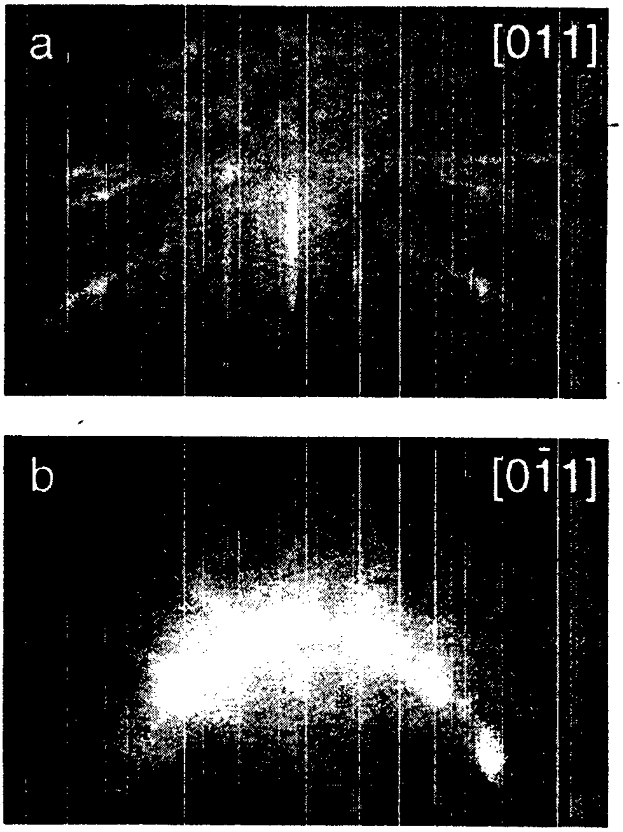


FIG. 1. RHEED patterns of a bare vicinal Si(100) 2×1 surface at 50 °C in the (a) [011] and (b) $[0\bar{1}1]$ azimuthal directions. A 1000 Å silicon epilayer was grown on the Si(100) substrate before it was annealed at 780 °C and then cooled to 50 °C.

While we obtained only one type of double-stepped bare vicinal Si(100) surface, we produced two types of double-stepped arsenic-covered vicinal Si(100):As surfaces with two different arsenic dimer orientations. The type depended on the surface treatment conditions. According to Kroemer's nomenclature,[12] a type A surface is composed of terraces with surface dangling bonds and/or surface dimer bonds parallel to the step edges, as schematically illustrated in Fig. 2(a), and a type B surface is composed of terraces with surface dangling bonds and/or surface dimer bonds perpendicular to the step edges, as schematically illustrated in Fig. 2(d). For arsenic-covered surfaces, a type A surface is a Si(100):As 2×1 reconstructed surface. The surface was prepared by initiating arsenic adsorption on a Si(100) 2×1 surface at a substrate temperature of 600 °C and an As_4 beam flux pressure of 1×10^{-5} mbar, followed by a decrease in substrate temperature to below 300 °C. A type B surface is a Si(100):As 1×2 reconstructed surface, which was prepared by initiating arsenic adsorption on a Si(100) 2×1 surface at room temperature followed by an increase in substrate temperature to 600 °C, all under an As_4 beam flux pressure of 1×10^{-5} mbar. These surface preparation procedures are consistent with those described by Bringans.[14] Smooth and single-domain surfaces were obtained as indicated by the streaky RHEED patterns in Figs. 2(b) and 2(c) for the Si(100):As 2×1 reconstructed surface, and Figs. 2(e) and 2(f) for the Si(100):As 1×2 reconstructed surface. We also

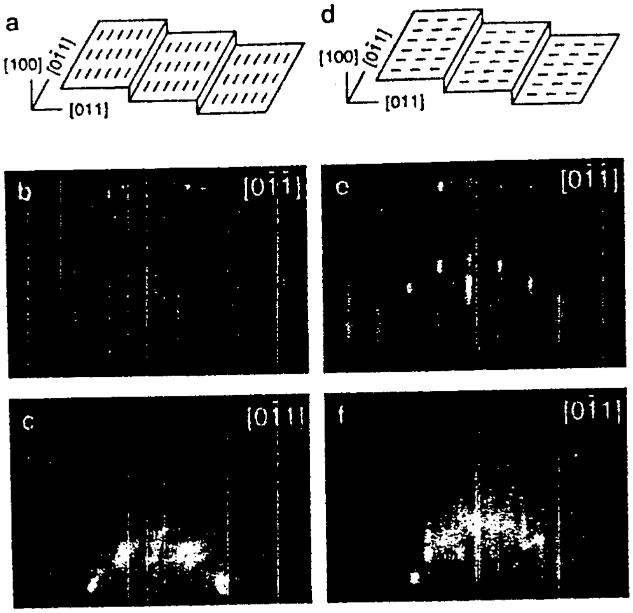


FIG. 2. (a) and (d) illustrate type A and type B terraces, with surface dimer bonds parallel and perpendicular to the step edges, respectively, of a vicinal Si(100) surface tilted towards the [011] direction. (b) and (c) show the RHEED patterns of a Si(100):As 2×1 reconstructed surface and (e) and (f) show the RHEED patterns of Si(100):As 1×2 reconstructed surfaces. [(b) and (e)], and [(c) and (f)] are the views at [0$\bar{1}\bar{1}$] and [0$\bar{1}$1] azimuthal directions, respectively.

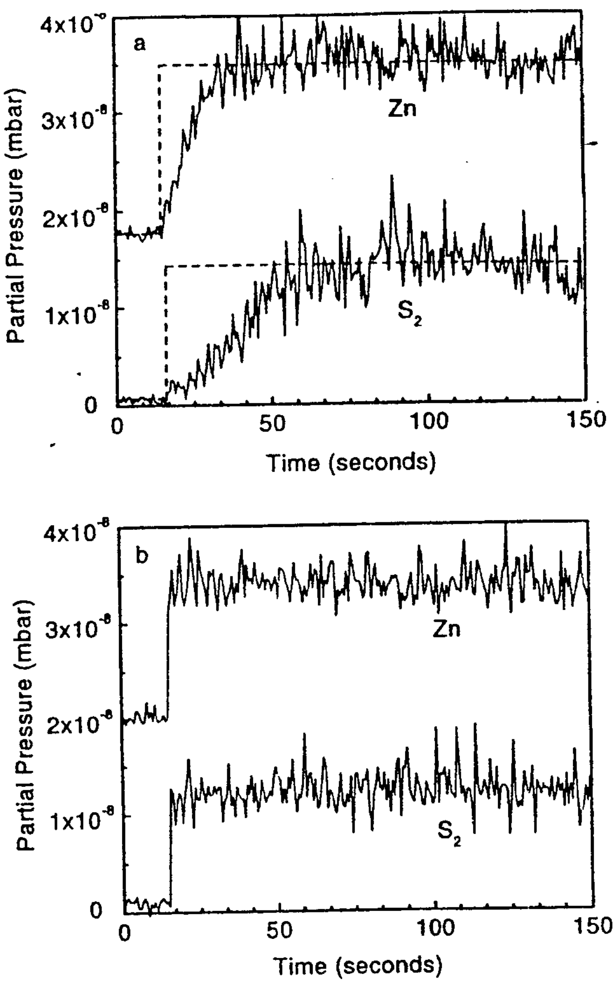


FIG. 3. Backscattered S_2 and Zn from (a) a bare Si(100) 2×1 surface and (b) a Si(100):As 2×1 surface. The adsorption experiments were conducted at a substrate temperature of 340 °C and a S_2+Zn beam flux pressure of 3×10^{-8} mbar.

note in Fig. 1 that the double-stepped bare vicinal Si(100) substrate has a type A surface. For either a type A or type B arsenic-covered surface, there is only one chemically adsorbed As monolayer on the surface. It is energetically unfavorable to have more than one adsorbed monolayer.

B. Surface adsorption

Adsorption experiments were carried out to reveal the difference between the ZnS growth on bare and arsenic-covered Si(100) surfaces during the initial growing periods. This was done by measuring the backscattered and desorbed species from the substrate surfaces. A high substrate temperature of 340 °C was used in these adsorption experiments. From the thickness of a ZnS layer grown at this temperature, we determined that, at this temperature, the sticking coefficients of Zn and S_2 on a ZnS surface were less than 5%. Therefore, the Zn and S_2 adsorption, observed during the initial growing period, can be mainly attributed to adsorption on the surface of a silicon substrate. Figure 3(a) shows the backscattered Zn and S_2, which were the only detectable zinc and sulfur containing species measured (using RGA) from a bare Si(100) surface at the beginning of a growth. Dashed lines in Fig. 3(a) depict the backscattered and/or desorbed Zn and S_2 with a negligible initial adsorption on the surface. The areas between the dashed lines and the solid lines are, therefore, attributed to the initial adsorption of zinc and sulfur on the silicon surface. From the data we find high initial sticking coefficients for both sulfur and zinc on a bare Si(100) surface. Accurate determination of the initial sticking coefficients was difficult because of the presence of both zinc and sulfur, which may interact with each other on the surface. As the surface became covered with zinc and sulfur, the sticking coefficients of zinc and sulfur decreased significantly to below 5%. Integration of the areas indicated that the surface coverage of Zn+S was close to one monolayer with a nonstoichiometric sulfur-to-zinc ratio of 2.6. In view of the stoichiometric beam fluxes of sulfur and zinc from the compound ZnS source, the stronger adsorption of sulfur is obvious.

Previous studies have revealed that sulfur reacts with silicon surface atoms forming disordered SiS_x.[9] One of our objectives was to develop methods of preventing this reaction in order to epitaxially grow ZnS on the silicon surfaces. For this purpose, we studied arsenic passivation. This method was introduced by Bringans *et al.* to prevent selenium from reacting with silicon surfaces.[10] Figure 3(b) shows the desorption of Zn and S_2 from a Si(100):As 2×1 surface. The desorption of Zn and S_2 leveled off immediately after the ZnS shutter was opened. The initial adsorption and/or reaction of sulfur and zinc on the silicon surface was clearly suppressed by the arsenic overlayer. A similar result was observed on Si(100):As 1×2 surfaces. These results clearly demonstrate the chemical passivation effect of an arsenic overlayer on a Si(100) surface.

As the adsorption experiments proceeded, we followed the surface structural changes by using RHEED. On a bare Si(100) surface, the RHEED pattern transformed from a streaky 2×1 pattern into a faint 1×1 pattern. This suggests a certain degree of surface disordering, although an amorphous layer was not formed. As we will explain in the following section, zinc adsorption on a Si(100) surface did not cause surface disordering. Therefore, the deterioration of surface ordering was attributed to sulfur adsorption. On the arsenic-passivated Si(100) surfaces, we found no change in the RHEED patterns throughout the $Zn+S_2$ adsorption period. Indeed, at the adsorption condition, i.e., substrate temperature at 340 °C and a $Zn+S_2$ beam-equivalent pressure of 3×10^{-8} mbar, the ZnS growth rate on the arsenic-passivated Si(100) surfaces was essentially zero, suggesting zero chemical adsorption on the surfaces. This further demonstrates the effectiveness of an arsenic passivation overlayer on Si(100) surfaces.

C. Surface structural changes during initial growth

Initial growth of ZnS on bare and on arsenic-passivated vicinal Si(100) surfaces was studied with the use of RHEED. The growth was conducted at a rate of 1.5 monolayers (ML) per minute. We noted that at this low growth rate, the epitaxial growth could proceed at a substrate temperature as low as 0 °C on a well-prepared vicinal Si(100) surface, with or without arsenic passivation. Here, we report data recorded at a substrate temperature of 50 °C. By measuring the backscattered Zn and S_2 beam fluxes from the substrate surface, we found that at this low substrate temperature, the sticking coefficients of Zn and S_2 were close to unity. Therefore, it was possible to determine the surface coverage and/or layer thickness of ZnS precisely by recording the shutter opening time of the ZnS source. In our previous study, conducted at higher substrate temperatures of 220–340 °C, the definitive surface coverage of ZnS in the initial growth stage was difficult to determine because of the low and variable sticking coefficients.[7] In this paper, we define one monolayer of ZnS as composed of one atomic layer of zinc plus one atomic layer of sulfur.

Figures 4(a) and 4(b), and 4(c) and 4(d), show the $[01\bar{1}]$ RHEED patterns of ZnS layers grown on a bare Si(100) 2×1 and on an arsenic-passivated Si(100):As 2×1 surface, respectively. At 0.75 ML coverage, a ZnS layer grown on a bare silicon surface exhibited sharper diffraction patterns [Fig. 4(a)] than one grown on an arsenic-passivated silicon surface [Fig. 4(c)]. This suggested that the epitaxial condition at the interface was better during the initial growth stage of ZnS on the bare Si(100) 2×1 surface than it was for growth on a Si(100):As 2×1 surface. However, at 3 ML in thickness, a ZnS film grown on a bare Si(100) surface exhibited a smeared RHEED pattern [Fig. 4(b)] containing oblique lines. These oblique lines can be attributed to microtwins and stacking faults, as will be described in detail later. At the same thickness, a ZnS layer grown on an arsenic-passivated Si(100):As 2×1 surface exhibited brighter and sharper RHEED patterns without noticeable oblique lines. While not shown, the RHEED-pattern evolution of a ZnS layer grown on a Si(100):As 1×2 surface exhibited similar behavior to that for ZnS grown on a Si(100):As 2×1 surface.

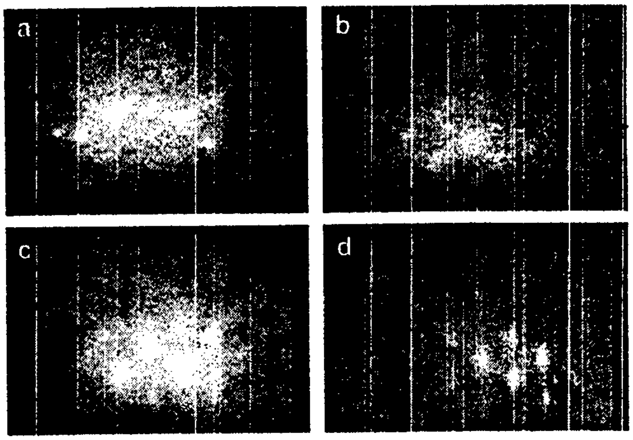


FIG. 4. RHEED patterns of ZnS surfaces at initiation of growth in the $[01\bar{1}]$ azimuthal direction. (a) 0.75 and (b) 3 ML of ZnS grown on a bare Si(100) 2×1 surface. (c) 1 ML and (d) 3 ML of ZnS grown on a Si(100):As 2×1 surface. The growth temperature was 50 °C and the growth rate was 0.025 ML per second.

As a test to find an alternative means of preventing chemical reactions between the sulfur and silicon surface, we studied the initiation of growth using zinc adsorption on the silicon substrate. A similar approach was used in a study of epitaxial growth of ZnSe on Si(100) substrates by Park and Mar.[15] Although the observation was left unexplained, these workers found that the initial exposure of a silicon substrate to a zinc flux was essential for achieving epitaxial growth of ZnSe layers. In our experiment, an ordered Si(100):Zn 3×1 surface structure was formed, as shown by the RHEED pattern of Fig. 5, after zinc was deposited on a well-prepared vicinal Si(100):As 2×1 surface. However, the maximum surface coverage of the ordered zinc was less than 0.5 ML over a substrate temperature range of 200–330 °C, as determined by temperature-programed desorption measurements.[7] At higher substrate temperatures, the sticking coefficient of zinc was too low to produce significant surface coverage. At

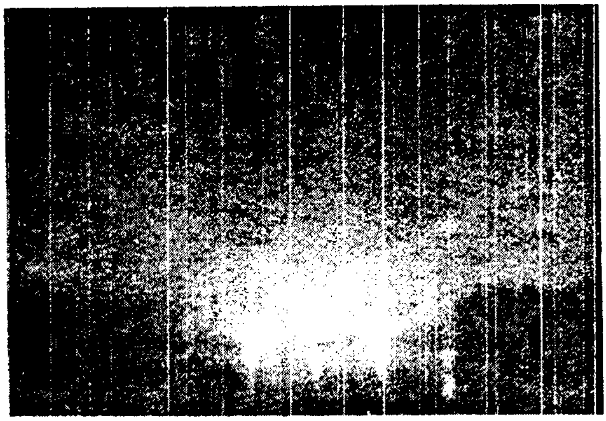


FIG. 5. A RHEED pattern of a Si(100):Zn 3×1 reconstructed surface at the [011] azimuthal direction. Zinc was deposited on a (vicinal) Si(100) 2×1 surface at a substrate temperature of 300 °C and a zinc beam flux pressure of 1×10^{-7} mbar.

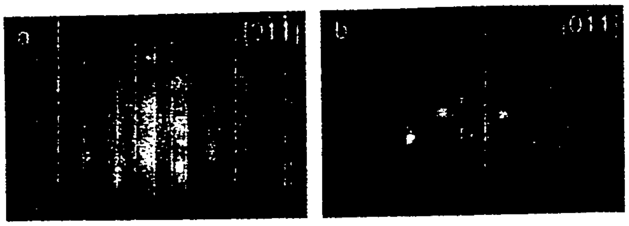


FIG. 6. RHEED pattern of a 25 Å ZnS layer grown on a Si(100) 2×1 surface at a substrate temperature of 50 °C, with (a) from $[0\bar{1}\bar{1}]$ and (b) from $[0\bar{1}1]$ azimuthal directions.

lower substrate temperatures, polycrystalline zinc formed. In addition, the growth of ZnS on a zinc-covered surface showed similar sulfur adsorption characteristics as shown in Fig. 3(a). Moreover, the RHEED patterns of the ZnS surface did not show significant improvement over the direct growth of ZnS on a bare Si(100) 2×1 surface. These results showed that the preadsorption of zinc on a Si(100) surface did not prevent the chemical adsorption of sulfur on the silicon surface and, hence, did not improve the epitaxial growth. No further study of this approach was pursued. It should be pointed out that our results were obtained with the use of a ZnS-compound source; whereas, the Park and Mar experiment was performed using separate elemental Zn and Se sources. It is likely that in our experiment, the reaction of sulfur with the silicon surface was already weakened by the simultaneous presence of zinc and, therefore, additional zinc did not produce any significant effect.

We also studied the adsorption of zinc on an arsenic-passivated silicon surface, as an attempt to improve the initial growth mode. However, in this case zinc (without the presence of sulfur) could not form an ordered single-phased layer. At substrate temperatures above 200 °C and a Zn beam-equivalent pressure of 1.5×10^{-7} mbar, the sticking coefficient of Zn was too low to obtain significant Zn surface coverage. At substrate temperatures below 200 °C, polycrystalline zinc films of a highly three-dimensional nature formed. Clearly, the energetically stable arsenic dimer bonds of a reconstructed Si(100):As surface prevented the formation of As–Zn bonds. Therefore, on an arsenic-passivated Si(100):As surface, zinc atoms alone cannot initiate epitaxial growth. In the discussion section below, we provide an explanation as to how epitaxial growth was initiated on an arsenic-passivated Si(100) surface in the presence of both zinc and sulfur.

D. Surface and crystal structure of ZnS layers grown on bare Si(100) surfaces

Figure 6 shows the RHEED patterns of a 25 Å thick ZnS film grown on a bare Si(100) 2×1 surface at a substrate temperature of 50 °C, annealed up to 300 °C for 5 min, and then cooled back to 50 °C. Undulating streaks were observed in the RHEED pattern along the $[0\bar{1}\bar{1}]$ azimuthal direction, as shown in Fig. 6(a). Elongated spots with oblique lines were observed in the RHEED pattern along the [011] azimuthal direction, as shown in Fig. 6(b). No surface reconstruction was observed. These RHEED patterns indicate a moderately smooth but defective surface structure. The oblique lines are attributed to microtwins in the ZnS layer. In addition, we note that the surface morphology and the surface crystal structure, as revealed by RHEED, were not improved as the film grew thicker. The reason for the lack of a smoothing effect, as the films grew thicker, will become clear as we present TEM results in the following discussion. It is also worth mentioning that the ZnS growth at a lower initial temperature resulted in a smoother surface. For example, our previous study of ZnS growth at a 220 °C initial temperature resulted in much spottier RHEED patterns than the ones shown in Fig. 6.[7]

Figure 7 shows cross-sectional TEM micrographs of a ZnS layer grown on a bare Si(100) 2×1 surface. Figure 7(a) is a $(01\bar{1})$ cross section, which shows a side view of the staircase of the surface steps. The orientation of the substrate-surface steps is illustrated in Fig. 2(a). In this micrograph, we observe a high density of microtwins and stacking faults, marked as T. It is interesting to note that while there are two sets of {111} planes, $(1\bar{1}\bar{1})$ and (111), in which microtwins and stacking faults could form, we observe the crystal defects only in $(1\bar{1}\bar{1})$ planes. Similar anisotropic characteristics of stacking faults were observed in GaAs epitaxial layers grown on the same type of vicinal Si(100) substrates.[16] By using TEM, we observed that the microtwins and stacking faults extend throughout the ZnS layer, with virtually a constant density. This explains why the ZnS surface structure and the smoothness were not improved as the layer grew thicker, as we described earlier. By counting the number of microtwins/stacking faults along the ZnS/Si interface and dividing this by the length of the interface, we determined the one-dimensional microtwin/stacking-fault density to be 2.6×10^{6} cm^{-1}. This value is the same as the density of surface double steps, along the [011] direction, on a vicinal Si(100) substrate of 4° offcut in the [011] direction, which was used in this study. This suggests a correlation between the substrate-surface steps and the formation of the microtwins and stacking faults.

Figure 7(b) is a high-resolution $(01\bar{1})$ cross-sectional TEM micrograph showing the interface region of the same ZnS/Si structure as that shown in Fig. 7(a). A white line was drawn along the interface as a visual guide. Although this particular micrograph does not contain a good atomic image for the silicon lattice, we can estimate the silicon surface morphology by following the interface line. From this micrograph, we learn that the silicon surface is composed of staircase-shaped terraces that are separated by double steps. Microtwins in the ZnS layer are initiated at step edges of the silicon surface and they are mostly 3 ML thick. In the discussion section, we propose a model to explain the formation of the microtwins at the surface-step edges.

Figure 7(c) is a (011) cross-section micrograph, which shows a front view of the staircase of the surface steps. This micrograph further reveals additional information about the structure of the microtwins and stacking faults in the ZnS layer. In this micrograph, we observe almost no microtwins and stacking faults lying in the {111} planes, $(11\bar{1})$ and $(\overline{111})$, that are perpendicular to the cross-section surface. However, we observe vertical lines, as indicated by the arrow D. We attribute the vertical lines to Shockley partial dislocations

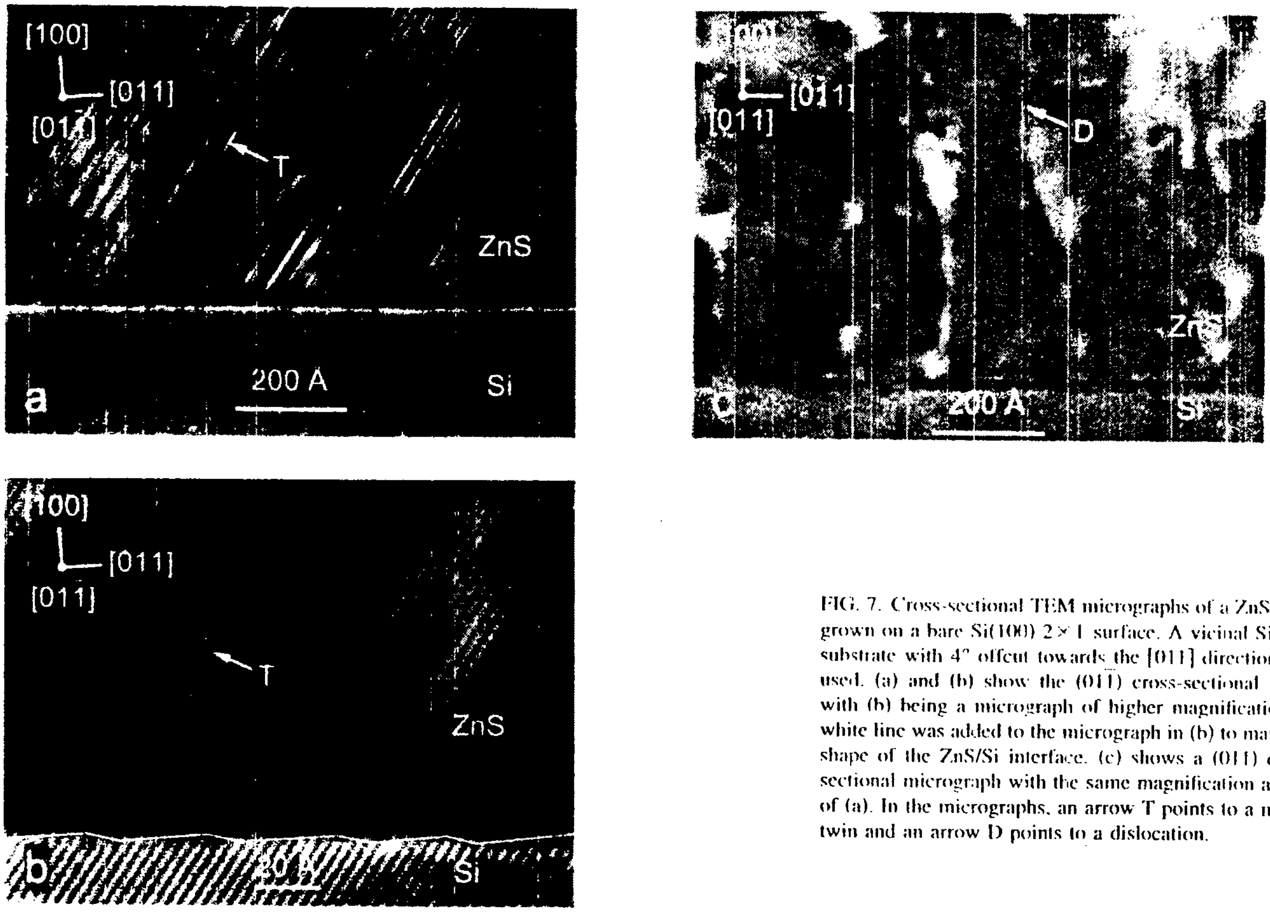


FIG. 7. Cross-sectional TEM micrographs of a ZnS layer grown on a bare Si(100) 2×1 surface. A vicinal Si(100) substrate with 4° offcut towards the [011] direction was used. (a) and (b) show the $(01\bar{1})$ cross-sectional views with (b) being a micrograph of higher magnification. A white line was added to the micrograph in (b) to mark the shape of the ZnS/Si interface. (c) shows a (011) cross-sectional micrograph with the same magnification as that of (a). In the micrographs, an arrow T points to a microtwin and an arrow D points to a dislocation.

connecting the stacking faults in the $(1\bar{1}\bar{1})$ planes. It is interesting to note that the stacking faults in the ZnSe layers grown on GaAs(100) substrates are mostly bound by Frank partial dislocations that appear as V shapes.[17] Later, we will discuss the geometric structures of the microtwins and stacking faults as derived from the micrographs shown in Fig. 7.

E. Structure of ZnS layers grown on arsenic-passivated Si(100) surfaces

Figures 8 shows RHEED patterns of two 25 Å thick ZnS films grown on arsenic-passivated Si(100):As 2×1 [Figs. 8(a) and 8(b)] and Si(100):As 1×2 [Figs. 8(c) and 8(d)] surfaces, respectively. The films were grown at a substrate temperature of 50 °C, annealed up to 300 °C for 5 min, and then cooled back to 50 °C. In Fig. 8(a), a RHEED pattern from the $[0\bar{1}\bar{1}]$ azimuthal direction, we observe half-order lines that are due to 2× reconstruction. In this image, the half-order lines are fuzzy because the ZnS surface was relatively rough, however, the half-order lines were unambiguously identified when we rotated the substrate to observe a RHEED pattern slightly off the $[0\bar{1}\bar{1}]$ azimuthal direction.[3] In Fig. 8(b), a RHEED pattern from the $[0\bar{1}1]$ azimuthal direction, we observe only integer-order lines. Combining Figs. 8(a) and 8(b), we identify the ZnS film grown on a Si(100):As 2×1 substrate surface as having a 2×1 reconstructed surface. Similarly, from Figs. 8(c) and 8(d), we observe a 1×2 reconstructed ZnS surface for a ZnS layer grown on a Si(100):As 1×2 substrate surface. At the low substrate temperature of 50 °C used in this study, the ZnS surfaces were sulfur stabilized and the surface reconstruction was attributed to the formation of sulfur dimers.[18] Therefore, our RHEED results indicated that surface S–S dimers were parallel to the initial As–As dimers. Besides the surface reconstruction, in Fig. 8(b), we notice a slight sign of oblique line formation that is similar to but less pronounced than that observed in Fig. 6(b) for a ZnS layer grown on a bare

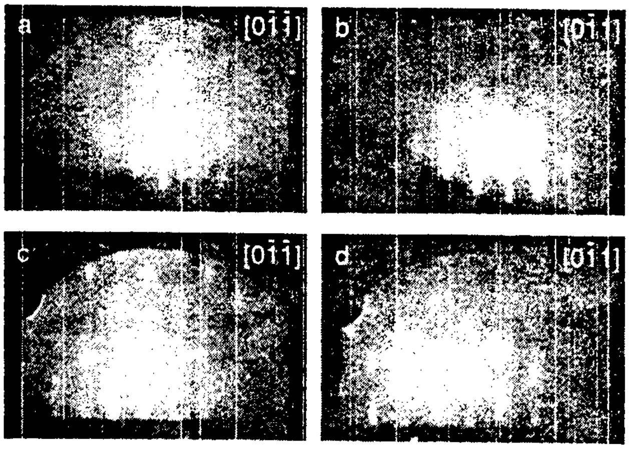


FIG. 8. RHEED patterns of 25 Å ZnS layers grown [(a) and (b)] on Si(100):As 2×1 and [(c) and (d)] Si(100):As 1×2 reconstructed surfaces at a substrate temperature of 50 °C. [(a) and (c)] and [(b) and (d)] are the views at $[0\bar{1}\bar{1}]$ and [011] azimuthal directions, respectively.

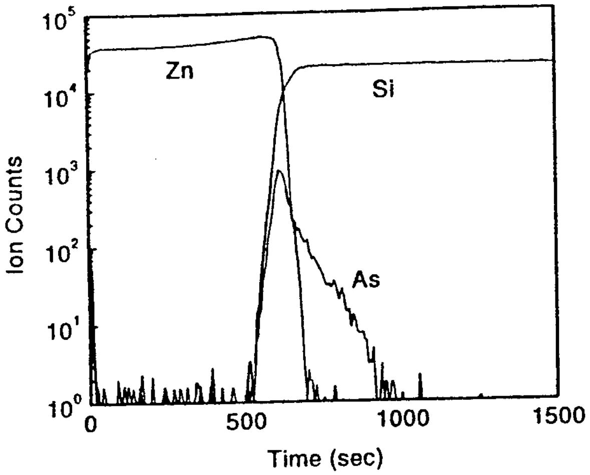


FIG. 9. A SIMS profile of a ZnS/As/Si(100) heterostructure.

$Si(100)\ 2\times 1$ substrate surface. These oblique lines are attributed to the presence of microtwins in the ZnS layer, as we will subsequently reveal by means of TEM micrographs.

A SIMS profile analysis was performed to determine whether the arsenic layer remained at the ZnS/Si interface region after ZnS growth or whether it floated on top of the ZnS surface as in the case of GeSi growth on arsenic-covered silicon surfaces.[19] The sample was composed of a 1500 Å ZnS layer grown on an arsenic-passivated Si(100):As 1×2 surface with an initial growth temperature of 220 °C for the first 50 Å and an elevated growth temperature of 300 °C for the rest of the layer thickness. Figure 9 shows the chemical profiles of Zn, Si, and As for this layer. The profile of As indicates that it remained at the interface region after the ZnS growth. In contrast, during the GeSi epitaxial growth on an arsenic-covered Si(100) substrate, arsenic floats on top of the GeSi surface and improves the surface morphology.[19] This effect, the so-called surfactant effect, is attributed to the reduction of surface free energy on an arsenic-covered GeSi surface as compared with a bare GeSi surface. The SIMS result of this study suggests that the function of arsenic in ZnS growth is different from that in GeSi epitaxial growth on arsenic-covered silicon substrates. Later, we will discuss the function of arsenic as an interfacial surfactant at the ZnS/Si(100) interface. In Fig. 9, we observe a high arsenic concentration in the silicon epitaxial layer (about 1000 Å). It is likely that the arsenic was incorporated into the silicon during the growth of a silicon buffer layer because of the high arsenic background pressure of $10^{-9}-10^{-8}$ mbar in our single-growth-chamber system. The use of a multi-growth-chamber system would significantly lower this background doping or cross-contamination effect.

TEM was used to study the effects of arsenic passivation on the crystal structure of the ZnS layers. Shown in Fig. 10 are TEM micrographs of a ZnS layer grown on an arsenic-passivated Si(100):As 1×2 surface. Comparing the crystal structure of the ZnS layer grown on the arsenic-passivated silicon surface, as shown in Fig. 10, with that of a ZnS layer grown on a bare silicon surface, as shown in Fig. 7, we note several major differences. As mentioned earlier, from Figs. 7(a) and 7(b), the $(01\bar{1})$ cross-sectional TEM micrographs of a ZnS layer grown on a bare vicinal $Si(100)\ 2\times 1$ surface, we observe that the crystal defects are mostly in the form of microtwins and only a few stacking faults exist. The microtwins/stacking faults are initiated at step edges of the substrate surface. The microtwins and stacking faults are almost exclusively in $(1\bar{1}\bar{1})$ planes. The one-dimensional crystal-defect density is as high as $2.6\times 10^6\ cm^{-1}$. By comparison, from Fig. 10(a), the $(01\bar{1})$ cross-sectional TEM micrographs of a ZnS layer grown on a Si(100):As 1×2 surface, we observe both microtwins and stacking faults with most of the crystal defects being stacking faults. Most of the crystal defects are not initiated at the silicon substrate surface. The microtwinning and stacking-fault defects are found in both (111) and $(1\bar{1}\bar{1})$ planes. The density of the microtwins and stacking faults is about 1/5 of that in the ZnS layer grown directly on a bare silicon surface. More differences are revealed by comparing the (011) cross-section micrographs of Figs. 7(c) and 10(b). While vertical lines due to Shockley partial dislocations are observed in Fig. 7(c) for the ZnS layer grown on a bare $Si(100)\ 2\times 1$ surface, no vertical line is observed in Fig. 10(b) for the ZnS layer grown on a Si(100):As 1×2 surface. Instead in Fig. 10(b), we observe

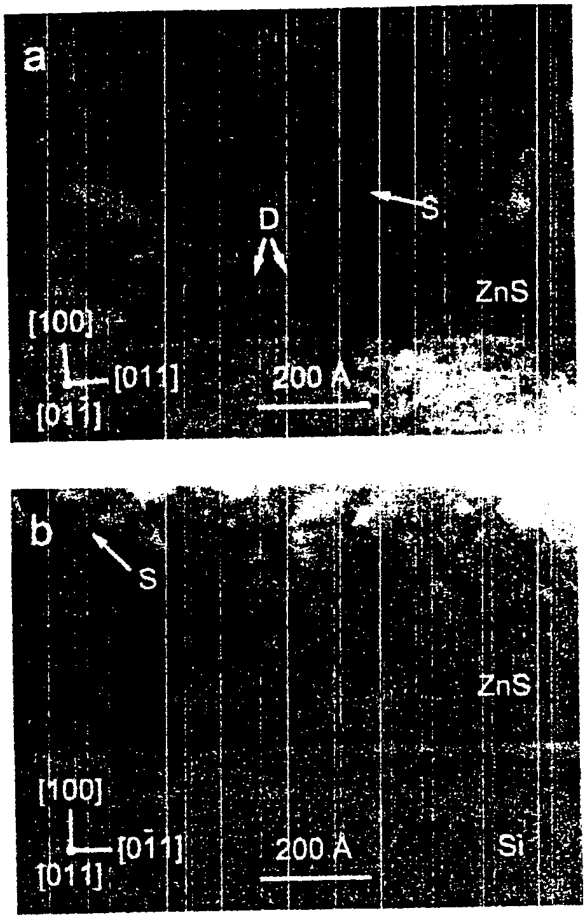


FIG. 10. Cross-sectional TEM micrographs of ZnS grown on a Si(100):As 1×2 surface at an initial substrate temperature of 50 °C for the first 25 Å and at 300 °C for the remaining 1500 Å. (a) shows a $(01\bar{1})$ cross-section view and (b) shows a (011) view. A Si(100) substrate with 4° offcut towards [011] direction was used.

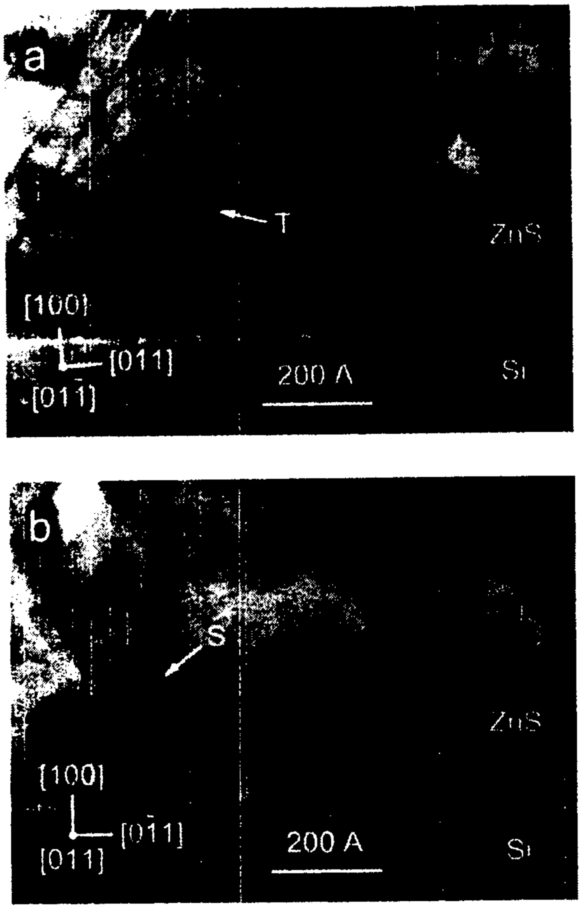


FIG. 11. Cross-sectional TEM micrographs of ZnS grown on a Si(100):As 2×1 surface at an initial substrate temperature of 50 °C for the first 25 Å and at 300 °C for the remaining 1500 Å. (a) shows a $(01\bar{1})$ cross-section view and (b) shows a (011) view. A Si(100) substrate with 4° offcut towards [011] direction was used.

stacking faults in the $(11\bar{1})$ and $(1\bar{1}1)$ planes, which are not observed in Fig. 7(c). These results suggest different initiation mechanisms for the formation of microtwins and stacking faults during the growth of ZnS layers on bare Si(100) 2×1 and on arsenic-covered Si(100):As 1×2 surfaces. Other features, observed in Fig. 10(a), are dislocations as indicated by arrow D. These are believed to be due to Frank partial dislocations connecting the stacking faults lying in $(11\bar{1})$ and $(1\bar{1}1)$ planes.

Figure 11 shows TEM micrographs of a ZnS layer grown on a Si(100):As 2×1 surface at the same growth conditions as those used for the sample in Fig. 10. The $(01\bar{1})$ cross-section view of Fig. 11(a) shows similar crystal-defect characteristics as that shown in Fig. 7(a) for the same $(01\bar{1})$ cross-section view of a ZnS layer grown on a bare Si(100) 2×1 surface. The crystal defects, observed in Fig. 11(a), are mostly microtwins, which originated on the substrate surface. These microtwins exhibit a strong anisotropic characteristic, existing mostly in $(1\bar{1}\bar{1})$ planes. However, the density of the microtwins and stacking faults was decreased to half of that in the ZnS layer grown on a bare Si(100) 2×1 surface.

Figure 11(b) shows a (011) cross-section micrograph of the same ZnS layer shown in Fig. 11(a). Unlike the $(01\bar{1})$ cross-section view, this micrograph is very different from the same cross-section view of the ZnS layer grown on a bare Si(100) 2×1 surface as shown in Fig. 7(c). No vertical line is observed in this micrograph. The image, however, is rather similar to Fig. 10(b), which shows the same (011) cross-section view of a ZnS layer grown on a Si(100):As 1×2 surface. In Fig. 11(b), we observe crystal defects in $(1\bar{1}1)$ and $(11\bar{1})$ planes. While rigorous assignment of these defects requires more detailed TEM analysis, some of these defects can be identified as stacking faults, as indicated by arrow S. Some other defects may be attributed to Frank partial dislocations that bound microtwins and stacking faults in (111) and $(1\bar{1}\bar{1})$ planes.

Last, we mention that the densities of microtwins and stacking faults in the ZnS layers grown on both Si(100):As 2×1 and Si(100):As 1×2 surfaces were found to depend on growth and substrate preparation conditions. For example, the growth of a silicon buffer layer to achieve a smooth substrate surface was found to be important in reducing defect densities. The preparation procedures for the deposition of an arsenic layer on vicinal Si(100) surfaces had a significant effect on the surface structure of the arsenic layer and, consequently, affected the defect densities of the ZnS layers.

IV. DISCUSSION

A. ZnS growth on bare silicon surfaces

We showed, in Fig. 3, that while both sulfur and zinc are strongly adsorbed on a bare Si(100) surface, the uptake of sulfur is about 2.6 times that of zinc in the initial growth period. In addition, we found no change in the adsorption characteristics of sulfur with or without the preadsorption of zinc. This suggests that the ZnS/Si(100) interface is mostly formed between sulfur and silicon atoms. The strong chemical adsorption and the reaction of sulfur on the Si(100) surface were studied in detail by Weser *et al.*,[9] who found the formation of disordered SiS_x ($x=1-4$) species at a substrate temperature above 200 °C. In a study of ZnSe growth on Si(100) surfaces, Bringans *et al.*[10] demonstrated the formation of amorphous ZnSe layers on the silicon surfaces and attributed this to the formation of disordered $SiSe_2$ during initial growth. We also found evidence for a chemical reaction between the sulfur and silicon surface, as described in our previous report.[7] By using a temperature-programed-desorption experiment, we observed the desorption of SiS from a ZnS-covered silicon surface at a substrate temperature of 430 °C. However, in our studies, we did not observe the formation of amorphous ZnS even at an initial growth temperature as high as 220 °C. We attributed this to the simultaneous supply of stoichiometric beam fluxes of zinc and sulfur from the ZnS-compound source that was used for the ZnS deposition. The reaction of sulfur and zinc competed with the reaction of sulfur and silicon. Therefore, the reaction of sulfur and silicon was limited.

In spite of the fact that ZnS and Si are closely lattice matched and that double-stepped vicinal Si(100) substrates were used, a high density of crystal defects were present in

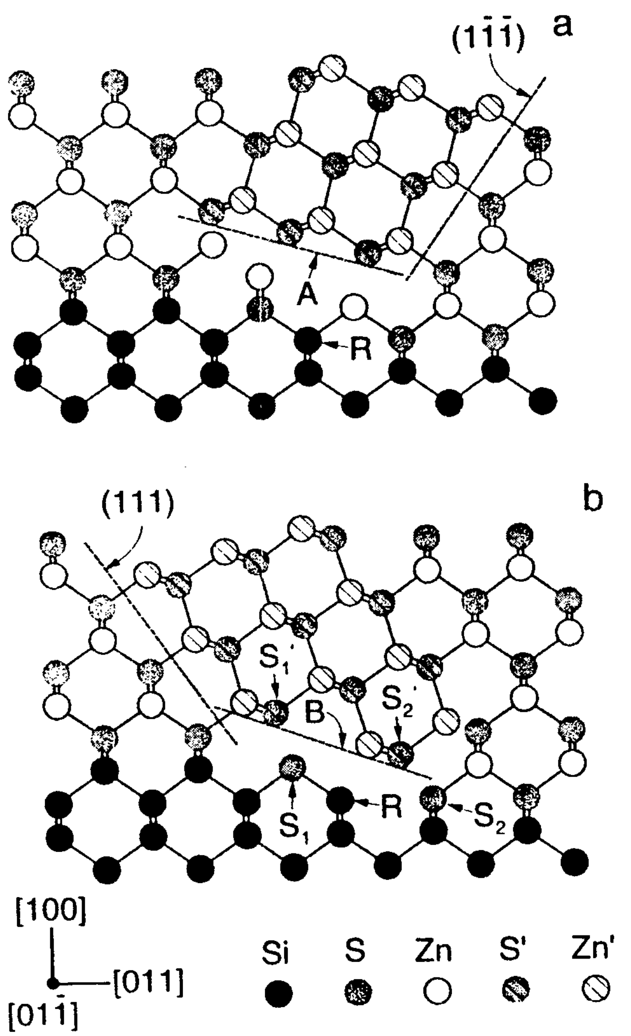


FIG. 12. Stick-and-ball models showing a (a) microtwinning defect having a $(1\bar{1}\bar{1})$ mirror plane and (b) microtwinning defect having a (111) mirror plane. Both defects are initiated at short-single steps formed by the rebonding silicon atoms *R*. Model (a) represents the experimentally observed microtwinning defect while model (b) represents the experimentally not observed microtwinning defect. In the figures, dashed line A in (a) marks a (111) facet, dashed line B in (b) marks a (011) facet, S′ and Zn′ mark the sulfur and zinc atoms that form microtwins.

the ZnS epitaxial layers grown directly on the silicon substrates, as revealed by the TEM micrographs shown in Fig. 7. As we described earlier, the crystal defects were mainly in the form of microtwins, although some stacking faults were also found. These microtwins formed only in $(1\bar{1}\bar{1})$ planes, i.e., one of four {111} planes. We have also presented evidence showing that the microtwins were initiated at the edges of surface steps.

In order to understand the formation mechanism of these microtwinning defects in the ZnS layers, we have constructed stick-and-ball atomic models of the ZnS/Si(100) interface in a surface-step region. A two-dimensional projection of the model is shown in Fig. 12. In our model, we assume that the silicon surface is fully covered by sulfur atoms and, therefore, has no dangling silicon bonds. This assumption is supported by our observations of strong sulfur adsorption on the silicon surface, as described earlier, and a low interfacial-state density of mid-10^{10} cm^{-2} eV^{-1}, as determined by capacitance–voltage measurements.[20]

Figure 12(a) shows a microtwinning defect having a $(1\bar{1}\bar{1})$ mirror plane, which represents the crystal defect observed in the TEM micrographs of Figs. 7(a) and 7(b). The step-edge structure of a vicinal Si(100) surface, was constructed based on a rebonded-atomic configuration, developed by Chadi.[21] According to Chadi, the most energetically stable structure of a double-step edge contains an extra row of rebonded silicon atoms. This rebonded-step model was confirmed experimentally in scanning tunneling microscope (STM) studies.[22] In our model we assume, after deposition of ZnS, that the rebonded silicon atoms undergo rearrangement and form a short single step as marked by R in Fig. 12. We attribute the initiation of microtwins to the presence of single-step structures at the silicon-surface-step edges. As seen in Fig. 12(a), the formation of microtwins at step edges leads to an alternating Zn/S layer structure along the vicinal-surface contour, and thereby avoids the formation of antiphase defects. Microtwins form easily in ZnS because of the low stacking-fault energy of ZnS. We note that ZnS has a low stacking-fault energy of 5 meV/atom and forms both zinc-blende and wurtzite structures.[23] By comparison, a typical zinc-blende–stable material such as GaAs has a higher stacking-fault energy of 47 meV/atom.[23] In our construction of the stick-and-ball models, it was found that a microtwin consisting of three atomic layers can fit into a normal crystal lattice with little strain. This three-layer-twinning configuration is supported by our TEM observations that the majority of the microtwinning images are three to five atomic layers thick, as seen in the $(0\bar{1}1)$ cross-section micrograph of Fig. 7(a). The microtwinning images thicker than three atomic layers are attributed to superimposed projections of deeper lying zigzag-shaped microtwinning layers in TEM samples that are approximately 200 Å in thickness. On a vicinal Si(100) surface, kinks exist along the edges of surface steps. These kinks lead to the formation of zigzag-shaped microtwinning layers, as we will describe later.

Figure 12(b) shows a model of a microtwin having a (111) mirror plane. This model represents a crystal defect that is not observed in the TEM micrograph of Figs. 7(a) and 7(b). By comparing the models in Figs. 12(a) and 12(b), we hope to understand the anisotropic nature of the microtwins, as observed in Figs. 7(a) and 7(b). The difference between the two models lies at the ends of the microtwins in the vicinity of a silicon step edge. The ZnS microtwin layer lying in the $(1\bar{1}\bar{1})$ planes, shown in Fig. 12(a), has a {111} end facet, however, the ZnS microtwin layer lying in the (111) plane, shown in Fig. 12(b), has a {011} end facet. In general, the more densely packed {111} facet has a lower surface energy than the less densely packed {011} facet. In addition, in order to form the structure shown in Fig. 12(b), we must force the same type of atoms, S_1'–S_1 and S_2'–S_2 as shown in Fig. 12(b), to be located immediately next to each other. Given the ionic nature of a polar material, the structure shown in Fig. 12(b) is not expected to be stable, due to the repulsive forces between the same charges.

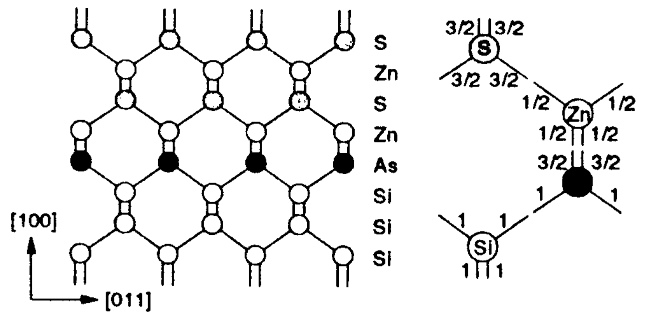


FIG. 13. Schematic illustration of a ZnS/As/Si(100) interface structure and electron distribution in the chemical bonds.

B. ZnS growth on arsenic-covered silicon surfaces

In this study, we have demonstrated improvement of the ZnS quality, such as better surface morphologies and reduced crystal-defect densities, by applying an arsenic layer on vicinal Si(100) substrate surfaces. Now we discuss the mechanisms for the improvement.

First, on a Si(100) surface, arsenic forms a 2×1 or 1×2 reconstructed structure, which is fully coordinated, highly stable, and, therefore, expected to have a passivating effect that inhibits surface chemical reactions.[14] The adsorption experiment of this study provided the first direct evidence of this passivation effect. The desorption/backscattering curves shown in Figs. 3(a) and 3(b) demonstrate a significant reduction in the initial adsorption of sulfur as well as zinc on an arsenic-passivated Si(100) surface as compared with that on a bare Si(100) surface. In addition, over a period of more than 5 min during the $Zn+S_2$ adsorption, we observed no change in the 2×1 RHEED pattern of the Si(100):As surface at substrate temperatures as high as 340 °C. This suggests a lack of significant sulfur reactions with the silicon surface. Therefore, the improvement in ZnS layer quality can be partially attributed to the chemical passivation effect of the arsenic layer, which provided a more stable initial epitaxial template.

The arsenic layer also acts as an "interfacial surfactant" at the interface of ZnS and Si on the (100) plane. Figure 13 shows a model of the ZnS/As/Si(100) interface structure. The model was first proposed by Bringans *et al.* based on an electron-counting argument.[10] According to this argument, the interface structure is chemically stable because all atoms in the interface region are fully coordinated with energetically stable octet–electron configurations. An illustration of the bonding-electron distribution appears in the right-hand side of Fig. 13. In the illustration, the arsenic interlayer links the two dissimilar substances, ZnS and Si; therefore, we call it an "interfacial surfactant." Without the arsenic interlayer, octet–electron configurations would not be possible at the ZnS–Si(100) interface. In a –Si–Si–Zn–S– structure, there is one electron too few in the Si and Zn interface atoms to fulfill the octet configuration. In a –Si–Si–S–Zn– structure, there is one electron too many in the Si and S interface atoms to fulfill the octet configuration. Therefore, neither structure is likely to be stable.

The experimental data of this study confirm the atomic model shown in Fig. 13. The SIMS profile data shown in Fig. 9 prove that the arsenic remains at the interface between ZnS and Si. The RHEED patterns in Fig. 8 demonstrate that the azimuthal orientation of a ZnS layer is determined by the azimuthal orientation of the initial arsenic layer on the silicon substrate. When the azimuth direction of the arsenic bonds was rotated by 90°, the azimuthal orientation of the ZnS layer followed. This orientation dependence is possible only when chemical bonds are formed between arsenic atoms and ZnS interface atoms. Our RHEED results also suggests that sulfur atoms in the ZnS lattice have the same bond orientation as that of interface arsenic atoms. From this orientation relation, we determined that the interface bonds are formed between zinc and arsenic atoms. Clearly, arsenic serves not only as a passivation layer in the initial growth period, but also as an interfacial agent to join the ZnS and Si surfaces, as explained earlier.

Although we argued that the ZnS/As/Si(100) interface structure is energetically stable, it is interesting to note that the initial ZnS growth on an arsenic-passivated Si(100):As surface was not in a layer-by-layer two-dimensional growth mode, as revealed by fading RHEED patterns shown in Fig. 4(c). The RHEED result suggests that within the first monolayer, the deposition of ZnS on the arsenic surface was somewhat disordered. Also, as we mentioned earlier, we did not observe the formation of an ordered phase of zinc chemisorbed on a Si(100):As surface. We attribute the observed phenomena to the highly stable 2×1 reconstructed Si(100):As surface on which chemical bonds with other atoms are difficult to form. As the surface coverage of zinc and sulfur increased above 1 ML, enhanced and sharper RHEED patterns appeared, indicating the formation of an ordered ZnS layer. The growth was indeed epitaxial as revealed by the RHEED patterns as shown in Fig. 8 and the high-resolution TEM micrographs as shown in Figs. 10 and 11. The formation of the As–Zn bonds is believed to be induced by the dipole–dipole interactions between Si–As and Zn–S dipoles.[24] This type of reaction sequence requires the presence of ZnS which, in turn, requires a certain critical surface population of zinc and sulfur atoms. This explains why crystallization occurred when the surface coverage of zinc and sulfur was increased beyond 1 ML.

The last subject to be discussed is the cause of the crystal-defect structure and density differences in ZnS layers grown on bare and on arsenic-passivated surfaces. While we have demonstrated a significant chemical passivation effect and suggested a modification of interfacial chemical bonding schemes by the use of an arsenic layer on Si(100) surfaces, these effects alone do not explain the differences. The surface structure of an arsenic layer must be important in determining the structure and density of ZnS defects, as significant differences between ZnS layers grown on Si(100):As 2×1 and Si(100):As 1×2 surfaces have been revealed by the TEM micrographs of Figs. 10 and 11.

The surface structure of the arsenic overlayer on vicinal Si(100) surfaces was studied by Bringans *et al.*[25] using STM. It was found that both Si(100):As 2×1 and Si(100):As 1×2 surfaces were composed of terraces of mostly single domains. However, a significant difference was found at the edges of the terraces or steps. On a Si(100):As 2×1 surface, the step edges were rough and contained closely spaced single steps in some cases. On a Si(100):As 1×2 surface, however, the step edges were smooth and had few single steps. We believe the closely spaced single steps on the Si(100):As 2×1 surface to be the cause of the microtwinning seen in the ZnS layers as shown by the TEM micrograph of Fig. 11(a). The situation is similar to ZnS growth on a bare Si(100), in which case, rebonding silicon atoms form single steps and cause the formation of microtwins. Therefore, the atomic model shown in Fig. 12(a) can be used to describe the formation of microtwins on the Si(100):As 2×1 surface, except that the interface sulfur atoms in the model are replaced by arsenic atoms. Although a ZnS layer grown on a Si(100):As 1×2 surface also contains stacking faults and few microtwins, as shown in the TEM micrograph of Fig. 10(a), few crystal defects are found to be initiated on the substrate surface. Obviously, the sharp double-stepped edges on the Si(100):As 1×2 surface are not the source of microtwins. Therefore, the decreased crystal-defect density of the ZnS layer grown on a Si(100):As 1×2 surface can be attributed to a better substrate template for the epitaxial growth.

The (011) cross-sectional TEM micrographs in Figs. 7(c), 10(b), and 11(b) provide additional information about the defect structures in all three ZnS layers. These micrographs suggest that the microtwins in the ZnS layer grown on a vicinal Si(100) 2×1 surface are bound by Shockley partial dislocations. However, the stacking faults and microtwins in the ZnS layer grown on vicinal Si(100):As 2×1 and Si(100):As 1×2 surfaces are bound by Frank partial dislocations. Based on this information and on our description of the microtwinning initiation at surface steps, we constructed models of microtwinning defects in ZnS layers grown on a bare vicinal Si(100) 2×1 surface and on a vicinal Si(100):As 2×1 surface, and schematically illustrate the models in Figs. 14(a) and 14(b), respectively. The stacking faults and microtwins in a ZnS layer grown on a Si(100):As 1×2 surface are not primarily initiated on the substrate surface. They are formed during the growth, probably by impurities, strains, and surface roughness. Bear in mind that ZnS has a low stacking-fault energy and, therefore, forms stacking faults and microtwins easily. Although a rigorous proof of the applicability of the crystal-defect models shown in Fig. 14 requires additional TEM analyses beyond those performed in the present study, all our observations about the structure of the crystal defects can be explained by these models. For example, these models would produce similar $(01\bar{1})$ cross-sectional micrographs for the ZnS layers grown on Si(100) 2×1 and Si(100):As 2×1 surfaces. The model in Fig. 14(a) would produce vertical lines, due to Shockley dislocations, in a (011) cross-sectional micrograph for a ZnS layer grown on a bare Si(100) surface, as shown in Fig. 7(c). The model in Fig. 14(b) would produce oblique lines, due to Frank dislocations, in a (011) cross-sectional micrograph for a ZnS layer grown on a Si(100):As 2×1 surface, as shown in Fig. 11(b). The model shown in Fig. 14(a) also explains the observation of microtwinning layers more than 3 ML thick in the $(01\bar{1})$ cross-sectional micrographs for the ZnS layers grown on Si(100) 2×1 surfaces; namely, the projection of zigzag-shaped microtwinning layers from thick TEM samples produce microtwinning images that appear to be more than 3 ML thick.

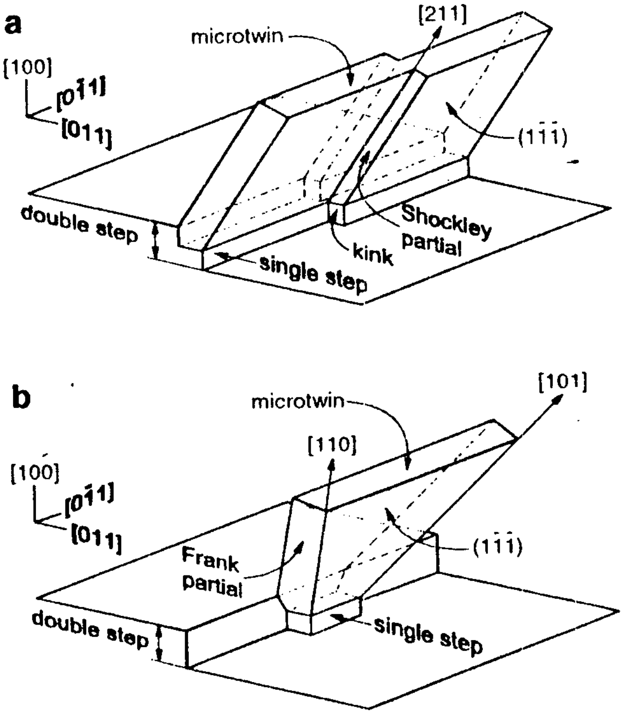


FIG. 14. Schematic illustration of the microtwinning defects in the ZnS layers grown on (a) a bare Si(100) surface and (b) a Si(100):As 2×1 surface.

V. CONCLUSIONS

This study has established a correlation between step-edge structures and the formation of microtwinning defects in the ZnS layers grown on vicinal Si(100) surfaces. Direct growth of ZnS on bare vicinal Si(100) 2×1 surfaces results in the formation of a high density of microtwinning defects. The initiation of the microtwinning defects is attributed to short single steps, formed by rebonding silicon atoms, at edges of double-step terraces on the vicinal surface.

A significant reduction in crystal defects, including stacking faults and microtwins, was found in ZnS layers grown on a Si(100):As 1×2 surfaces as compared to layers grown on bare Si(100) 2×1 surfaces. This effect is attributed to the elimination of single steps at the edges of the double-step terraces. The ultimate one-dimensional defect density was found to be about 5×10^5 cm^{-1}.

On the other hand, for ZnS layers grown on Si(100):As 2×1 surfaces, a relatively high density of microtwinning defects was found and the defects were initiated at the substrate surface. The seeds of the defects were identified with the short single steps at the edge of the double-step terraces of the vicinal Si(100):As 2×1 surface.

While the chemical reaction between sulfur and silicon is harmful to the epitaxial growth of ZnS, we did not find evidence that showed the crystal defects in the ZnS layers came directly from this chemical reaction. Because a ZnS compound source was used in this study, the presence of zinc is believed to have limited the chemical reaction between sulfur and silicon. More importantly, an arsenic overlayer at the interface was found to have a passivating effect, and thereby, inhibited the adsorption and reaction of sulfur and zinc atoms with a Si(100) surface. This interfacial overlayer of arsenic significantly improved the crystalline quality of ZnS films grown epitaxially on Si(100) surfaces.

ACKNOWLEDGMENTS

This work was supported in part under NSF Grant No. ECS-9306293, NASA Grant No. NAGW-1194, AFOSR/ARPA Grant No. F49620-96-C-0006, AFOSR Grant No. F49620-96-1-0242, Texas High Education Coordinating Board Grant No. ARP-999903-219, and TAMU/ERP Grant No. 95-54. The authors gratefully thank Robert Klima for his technical support and Robert T. Bate for useful discussions. The authors acknowledge helpful discussions with Dr. R. D. Bringans and Dr. L. T. Romano of Xerox Corp. The authors also acknowledge the generous assistance of Dr. M. Anthony, Dr. P. Chen, Dr. B. Gnade, and Dr. C. C. Cho, of Texas Instruments, in providing SIMS analysis, apparatus components, and helpful discussions. The authors acknowledge the Electron Microscopy Center of Texas A&M University for providing the TEM facilities.

[1] R. Houdré and H. Morkoç, CRC Crit. Rev. Solid State Mater. Sci. **16**, 91 (1990).

[2] X. Zhou, S. Jiang, F. Li, G. F. Spencer, R. T. Bate, and W. P. Kirk, in Proceedings of the 1995 IEEE Cornell Conference on Advanced Concepts in High Speed Semiconducting Devices and Circuits **1**, 498 (1995).

[3] X. Zhou, F. Li, G. F. Spencer, and W. P. Kirk, Mater. Res. Soc. Symp. Proc. **380**, 3 (1995).

[4] S. M. Sze, in *Physics of Semiconductor Devices* (Wiley, New York, 1981).

[5] C. Maierhofer, S. Kulkarni, M. Alonso, T. Reich, and K. Horn, J. Vac. Sci. Technol. B **9**, 2238 (1991).

[6] M. Yokoyama, K. I. Kashiro, and S. I. Ohta, J. Cryst. Growth **81**, 73 (1987).

[7] X. Zhou and W. P. Kirk, Mater. Res. Soc. Symp. Proc. **318**, 207 (1994).

[8] L. T. Romano, R. D. Bringans, X. Zhou, and W. P. Kirk, Phys. Rev. B **52**, 11 201 (1995).

[9] T. Weser, A. Bogen, B. Konrad, R. D. Schnell, C. A. Schug, and W. Steinmann, in *Proceedings of the Eighteenth International Conference on the Physics of Semiconductors*, Stockholm, Sweden, 1986, edited by O. Engström (World Scientific, Singapore, 1987), pp. 97–100.

[10] R. D. Bringans, D. K. Biegelsen, L.-E. Swartz, F. A. Ponce, and J. C. Tramontana, Phys. Rev. B **45**, 13 400 (1992).

[11] R. C. Henderson, J. Electrochem. Soc. **119**, 772 (1972).

[12] H. Króemer, Mater. Res. Soc. Symp. Proc. **67**, 3 (1986).

[13] K. Sakamoto, T. Sakamoto, K. Miki, and S. Nagao, J. Electrochem. Soc. **136**, 2705 (1989).

[14] R. D. Bringans, CRC Crit. Rev. Solid State Mater. Sci. **17**, 353 (1992).

[15] R. M. Park and H. A. Mar, Appl. Phys. Lett. **48**, 529 (1986).

[16] P. Lao, W. C. Tang, K. C. Rajkumar, S. Guha, and A. Madhukar, J. Appl. Phys. **67**, 6445 (1990).

[17] S. Guha, J. M. DePuydt, G. E. Hfler, M. A. Haase, B. J. Wu, and H. Cheng, Appl. Phys. Lett. **63**, 3023 (1993).

[18] J. M. DePuydt, H. Cheng, J. E. Potts, T. L. Smith, and S. K. Mohapatra, J. Appl. Phys. **62**, 4756 (1987).

[19] M. Copel, M. C. Ruter, E. Kaxiras, and R. M. Tromp, Phys. Rev. Lett. **63**, 632 (1989).

[20] (Unpublished).

[21] D. J. Chadi, Phys. Rev. Lett. **59**, 1691 (1987).

[22] J. E. Griffith, G. P. Kochanski, J. A. Kubby, and P. E. Wierenga, J. Vac. Sci. Technol. A **7**, 1914 (1989).

[23] S. Takcuchi, K. Suzuki, K. Maeda, and H. Iwanaga, Philos. Mag. A **50**, 171 (1984).

[24] Private communication with R. D. Bringans of Xerox Cooperation.

[25] R. D. Bringans, D. K. Biegelsen, and L.-E. Swartz, Phys. Rev. B **44**, 3054 (1991).

ELSEVIER

Journal of Crystal Growth 175/176 (1997) 624–631

JOURNAL OF CRYSTAL GROWTH

Molecular beam epitaxy of BeTe on vicinal Si(1 0 0) Surfaces

Xiaochuan Zhou[1], Shan Jiang, Wiley P. Kirk*

NanoFAB Center, Engineering-Physics Building, Texas A&M University, College Station, Texas 77843-4242, USA

Abstract

BeTe was epitaxially grown on vicinal Si(1 0 0) substrates for the first time. Reflection high energy electron diffraction (RHEED) was used to study the initial growth mode and the surface structure. Transmission electron microscopy (TEM) was used to study the crystal quality. The material was grown on both bare and arsenic covered vicinal Si(1 0 0) surfaces. Smooth and well-ordered surface structures readily formed during the growth. Low stacking-fault densities were obtained in the BeTe epilayers. The potential applications of the new material are discussed briefly.

PACS: 81.15.Hi; 81.05.Dz; 81.05.Cy

---

1. Introduction

The growth of compound semiconductors on silicon substrates has been extensively studied as a means of integrating silicon-based electronic devices with compound-semiconductor-based optoelectronic devices [1]. Such studies were recently extended to the development of silicon-based quantum devices, in which silicon and compound semiconductors are both involved in bandgap tailoring [2]. The growth of II–VI compound semiconductors on Si(1 0 0) surfaces is unique because the interface charge problem can be eliminated by the use of a monolayer of group III or group V elements at the interface [3]. The use of an arsenic passivation layer was introduced by Bringans et al. for the purpose of growing ZnSe on Si(1 0 0) substrates [4]. We later used the arsenic passivation method to grow ZnS layers for the purpose of developing ZnS/Si double-barrier resonant-tunneling diodes, in which ZnS was used as a wide-bandgap tunneling barrier material [5, 6]. While the crystallinity of the ZnS and ZnSe layers was significantly improved by use of the arsenic passivation method, the problem of high stacking-fault densities in both layers remained. This has been attributed to the intrinsically low stacking-fault energies of ZnS and ZnSe materials [3]. We were, therefore, prompted to search for new semiconductor materials that have wide bandgaps, could be grown epitaxially on silicon substrates, and have high stacking-fault energies.

---

* Corresponding author. Fax: + 1 409 845 2590; e-mail: kirk@nanofab.tamu.edu.

[1] Present address: Electronics Manufacturing Laboratory, University of Michigan, Ann Arbor, Michigan 48105-2551, USA.

0022-0248/97/$17.00 Copyright © 1997 Published by Elsevier Science B.V. All rights reserved

PII S0022-0248(96)00911-6

The result of our search has focused on a group of beryllium-VI compound semiconductors, including BeS, BeSe, and BeTe. According to the literature, BeS, BeSe, and BeTe, all have zincblende crystal structures, with lattice constants of 4.8630, 5.1477, and 5.6269 Å, and bandgaps of 6.10, 4.50, and 2.7 eV, respectively [7–9]. Using a simple linear interpolation, we find that an alloy, $BeSe_{0.45}Te_{0.55}$, can be lattice matched to silicon with a bandgap of 3.51 eV. Therefore, the Be-VI compound semiconductors should make good candidates for our study.

Among the most important features of the Be-VI compound semiconductors is their low ionicities. According to Phillips [10], this group of compounds have the lowest ionicities among all II–VI compound semiconductors. As a matter of fact, according to the same reference, their ionicities are even lower than that of GaAs. In general, there is a strong correlation between the stacking-fault energy and the ionicity of III–V and II–VI compounds. For example, Takeuchi et al. [11] developed an empirical correlation between the stacking fault energy and the charge redistribution index of both II–VI and III–V compounds. The charge redistribution index measures the charge transfer accompanying the strain of the bond length and can be related to the Phillips' ionicity according to Phillips and Van Vechten [12]. Using Takeuchi's method, we calculated the stacking fault energies of the Be-VI compound semiconductors and plotted the results in Fig. 1 along with the experimental data of other II–VI and III–V compound semiconductors. According to this prediction, all three Be-VI compounds have high stacking-fault energies, which are interestingly even higher than that of GaAs.

The study of Be-VI semiconductors may also have an impact on the improvement of II–VI blue lasers [13]. The short lifetime problem of the present II–VI blue lasers arises from the ease with which crystal defects generate and propagate in ZnSe based materials [14, 15] and this is believed to be the result of the weak ionic chemical bonds [16]. The source of the dislocations in the ZnSe-based laser structures has been identified with stacking faults that originate at the ZnSe/GaAs interface regions [14]. The lower ionicities, stronger chemical bonds, and higher stacking-fault energies of the Be-VI semiconductors may help to strengthen the II–VI laser materials and to increase the lifetimes of the laser devices. This concept is supported by a recent theoretical study of lattice elastic rigidity of II–VI semiconductor materials [17].

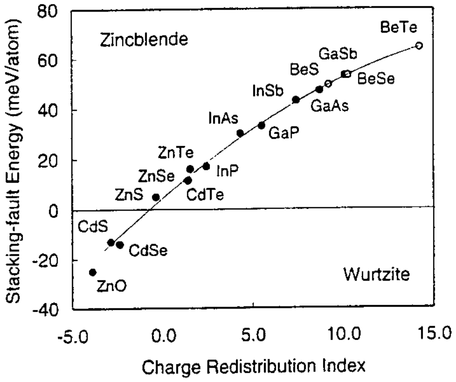


Fig. 1. Stacking fault energy of III–V and II–VI compounds. Solid dots are the experimental data, from Ref. [11], and open circles are the predicted values according to an empirical calculation method given in the same reference.

In this paper, we report initial results of molecular beam epitaxial growth and TEM studies of BeTe on vicinal Si(1 0 0) substrates.

2. Experimental procedure

Epitaxial growth was carried out in a Fisons VG-V80S MBE system. The system was equipped with an e-beam evaporator for silicon deposition and three effusion cells for Be, Te, and As sources. The MBE system was also equipped with a 15 keV RHEED system for surface structural studies and a Dycor M200 residual gas analyzer (RGA) for gas-phase composition analysis. The RGA was positioned in the line of sight of the substrate and had an orifice to confine the detection only to those atoms/molecules backscattered or desorbed from the substrate surface.

N-type vicinal Si(1 0 0) substrates with a 4° offcut towards [0 1 1] direction were used. The silicon substrates were chemically cleaned using a peroxide cleaning procedure, which left surfaces covered with a thin layer of oxide formed in a $HCl : H_2O_2 : H_2O = 3 : 1 : 1$ solution [18]. The substrates were annealed in the MBE preparation chamber at 870°C for 20 min and then transferred into the growth chamber. A 1000 Å thick silicon buffer layer was grown on each substrate at a temperature of 620°C to produce a smooth starting surface. Double stepped surfaces were obtained by annealing the substrate at 870°C and then gradually lowering the temperature to room temperature. RHEED patterns of single-domain 2 × 1 reconstructed Si(1 0 0) surfaces were observed.

Arsenic deposition on the silicon substrates was performed by using the procedures described by Bringans et al. [19]. Single-domain Si(1 0 0) : As 2 × 1 and Si(1 0 0) : As 1 × 2 reconstructed surfaces were obtained by depositing arsenic at an As_4-beam partial pressure of 1.5×10^{-5} mbar and at substrate temperatures of 570°C and 0°C, respectively.

BeTe layers were grown in two steps. The first 50–100 Å thick layer was grown at a substrate temperature of 400°C and the remaining layer was grown at 500°C. The growth rate was 0.2–0.4 μm/h, and the Te_2/Be beam partial-pressure ratio was about 10. Since this work represents an initial effort, the growth conditions have by no means been optimized. The layer thickness and the growth rate was determined by using TEM. It was found that BeTe was somewhat air sensitive. After an extended exposure of a BeTe film to air, the film darkened in color. In order to protect the BeTe surfaces from contacting air, we deposited a 100–300 Å thick silicon-cap layer on top of the BeTe surface. According to the observation on Be-VI bulk materials by Yim et al. [8], BeSe and BeS are stable in air. Therefore, the formation of a ternary compound, such as BeSeTe and BeSTe, is expected to improve the stability in air.

Cross-sectional TEM samples were prepared by standard mechanical cutting, lapping and ion-milling procedures. To reduce the time of reaction with air, after ion milling, the samples were immediately transferred into a TEM system. All TEM micrographs were taken by using a JEOL 2010 microscope at 200 kV.

3. Results and discussions

3.1. BeTe growth on a bare Si(1 0 0) surface

RHEED was used to study the growth process. During the BeTe growth on a bare Si(1 0 0) 2 × 1 surface, the first 100 Å of the growth was deposited in island-growth mode, as characterized by spotty RHEED patterns. Although the island-growth mode can be partially attributed to the lattice mismatch between BeTe and the silicon substrate, the main reason is the unfavorable interfacial chemistry. It is likely that the formation of an interface dipole between Si(1 0 0) and BeTe(1 0 0) prevented the formation of an abrupt interface [20]. This was demonstrated in our previous study of ZnS growth on Si(1 0 0) surfaces [5]. While ZnS and silicon are almost lattice-matched, the initial deposition was characterized as island growth.

As the BeTe film grew thicker, from about 100 to 150 Å up, streaky 2 × 1 RHEED patterns became apparent, indicating the formation of a smooth and well ordered surface. These RHEED patterns are not shown here because they are very similar to the ones shown in Fig. 2c and Fig. 2d for a BeTe layer grown on a Si(1 0 0) : As 2 × 1 surface. The BeTe growth was conducted under a tellurium over-pressure condition, therefore the BeTe surface was tellurium rich and the 2 × 1 RHEED pattern was attributed to the formation of Te–Te dimers. We noted that the BeTe surface and the initial Si(1 0 0) substrate surface had the same 2 × 1 reconstruction. This indicates that the surface tellurium dimers have the same azimuthal orientation as that of silicon dimers on the substrate surface, both being parallel to the surface-step edges. Since, in silicon and zincblende crystal structures along the ⟨1 0 0⟩ directions, atoms in every two atomic layers have the same bond orientations, the atomic sequence at the BeTe/Si(1 0 0) interface must be -Si–Si–Be–Te–. In other words, the first layer on the Si(1 0 0) surface was beryllium. This is different from the interface structures of ZnS/Si(1 0 0) and ZnSe/Si(1 0 0), grown by molecular beam epitaxy. In those cases, due to the strong chemical adsorption of sulfur and selenium (group VI elements) on Si(1 0 0) surfaces, the first layers were dominantly sulfur and selenium [5, 4]. It is interesting to note that during our initial growth of BeTe on a Si(1 0 0) surface, the tellurium shutter was actually opened first. The formation of Si–Be bonds implies a stronger adsorption of Be than Te on a Si(1 0 0) surface. A detailed study on the surface structure of Te adsorption on Si(1 0 0) 2 × 1 surfaces was performed by Higuchi et al. [21], who found that Te formed ordered phases up to one monolayer of coverage at 100°C. However, to our knowledge, no study of Be adsorption on Si(1 0 0) surfaces has yet performed.

3.2. BeTe growth on arsenic passivated Si(1 0 0) surfaces

On Si(1 0 0) : As 2 × 1 and Si(1 0 0) : As 1 × 2 reconstructed surfaces, BeTe growth exhibited a much improved initial growth mode, as compared with that on bare Si(1 0 0) 2 × 1 surfaces. Fig. 2(a), Fig. 2(b), and Fig. 3(a), Fig. 3(b) are the RHEED patterns of Si(1 0 0) : As 2 × 1 and Si(1 0 0) : As 1 × 2 surfaces, respectively. On a Si(1 0 0) : As 2 × 1 surface, the As–As dimers are parallel to the surface-step edges, while on a Si(1 0 0) : As 1 × 2 surface, the As–As dimers are perpendicular to the surface-step edges. On the two Si(1 0 0) : As surfaces, we observed similar BeTe growth behavior. During the first one or two monolayers of BeTe growth, the RHEED patterns became diffused and lost reconstruction lines. However, from the third monolayer (about 8 Å) up, the RHEED patterns became streaky and contained reconstruction lines, implying the formation of smooth and ordered BeTe surfaces. It is interesting to note that the BeTe growth on the Si(1 0 0) : As surfaces was neither in the Volmer–Weber growth mode (featured with initial three-dimensional growth) nor in the Stranski–Krastanov growth mode (featured with initial two-dimensional growth followed by the formation of three-dimensional islands). The diffused RHEED patterns at the very early stage of growth suggest that some initially adsorbed atoms do not occupy the exact epitaxial sites. However,

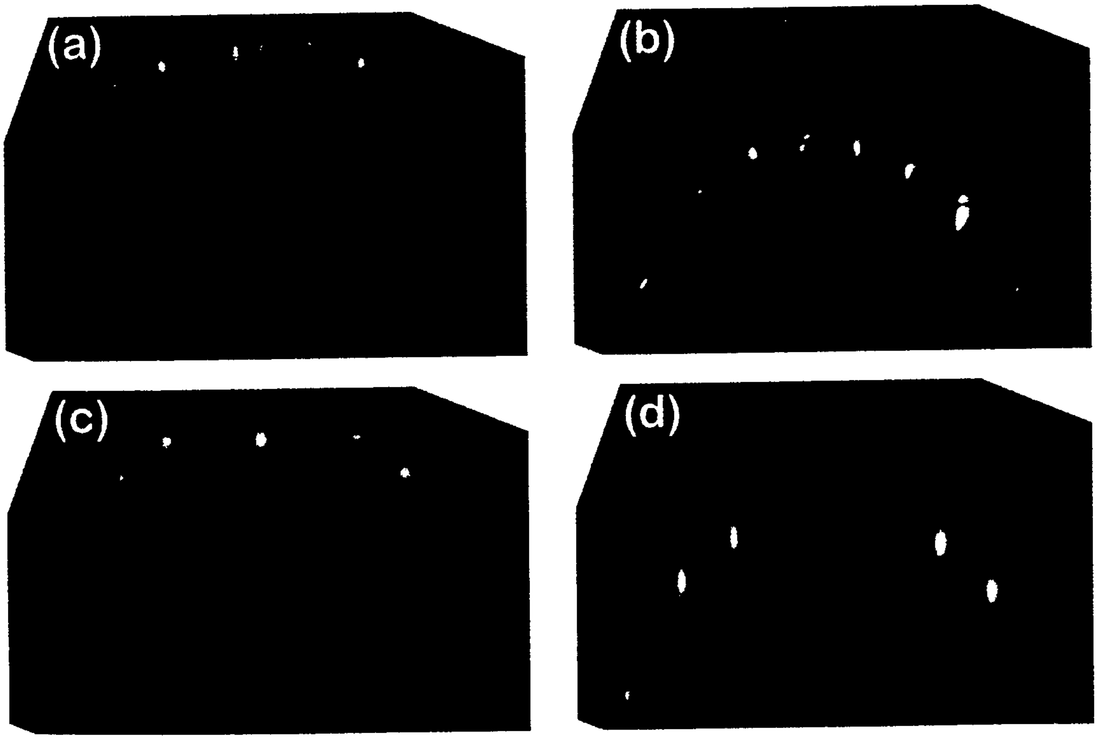


Fig. 2. RHEED patterns of a Si(1 0 0) : As 2 × 1 surface at (a) $[0\,\bar{1}\,\bar{1}]$ and (b) $[0\,\bar{1}\,\bar{1}]$ azimuthal directions, and of a BeTe 2 × 1 surface at (c) $[0\,\bar{1}\,\bar{1}]$ and (d) $[0\,\bar{1}\,\bar{1}]$ azimuthal directions.

the surface quickly became smooth and well ordered after adding just a couple of monolayers of BeTe to the surface, a similar phenomenon was observed in our study of ZnS growth on Si(1 0 0) : As surfaces. The exact growth mechanism is yet to be understood.

The layer thickness in the initial growth period, as given above, is an estimate, and was derived from the growth time and the average growth rate. However, by using a residual-gas analyzer in front of the substrate, we detected the desorption of beryllium from the substrate during the growth even at a low growth temperature of 400°C. This indicates that the sticking coefficient for Be is less than one. We also observed a higher beryllium desorption, as measured by the Be partial pressure, at the beginning of the growth. While the partial-pressure transient might be attributed to the temperature change in the beryllium effusion cell after the shutter opening, we leave open the possibility that the observation was due to a transient in the growth rate during the initial growth period. In any case, it is reasonable to expect a difference in the sticking coefficient of beryllium on a Si(1 0 0) : As surface as compared to a BeTe surface. Since we observed higher Be desorption in the initial growth period, the actual growth-mode transition from a somewhat disordered deposition into a planar epitaxial growth could have occurred at a BeTe thickness less than 3 monolayers.

Shown in Fig. 2c, Fig. 2d, Fig. 3c and Fig. 3d are the RHEED patterns of BeTe layers grown on the Si(1 0 0) : As 2 × 1 and Si(1 0 0) : As 1 × 2 surfaces, respectively. These RHEED patterns indicate the formation of BeTe 2 × 1 and BeTe 1 × 2 reconstructed surfaces on the two Si(1 0 0) : As surfaces. As mentioned earlier, the BeTe surfaces were terminated with tellurium dimers. For the same reason as explained earlier, we find that the BeTe layers have the same Te-bond orientations as those of initial arsenic atoms, regardless of starting with a Si(1 0 0) : As 2 × 1 or a Si(1 0 0) : As 1 × 2 surface.

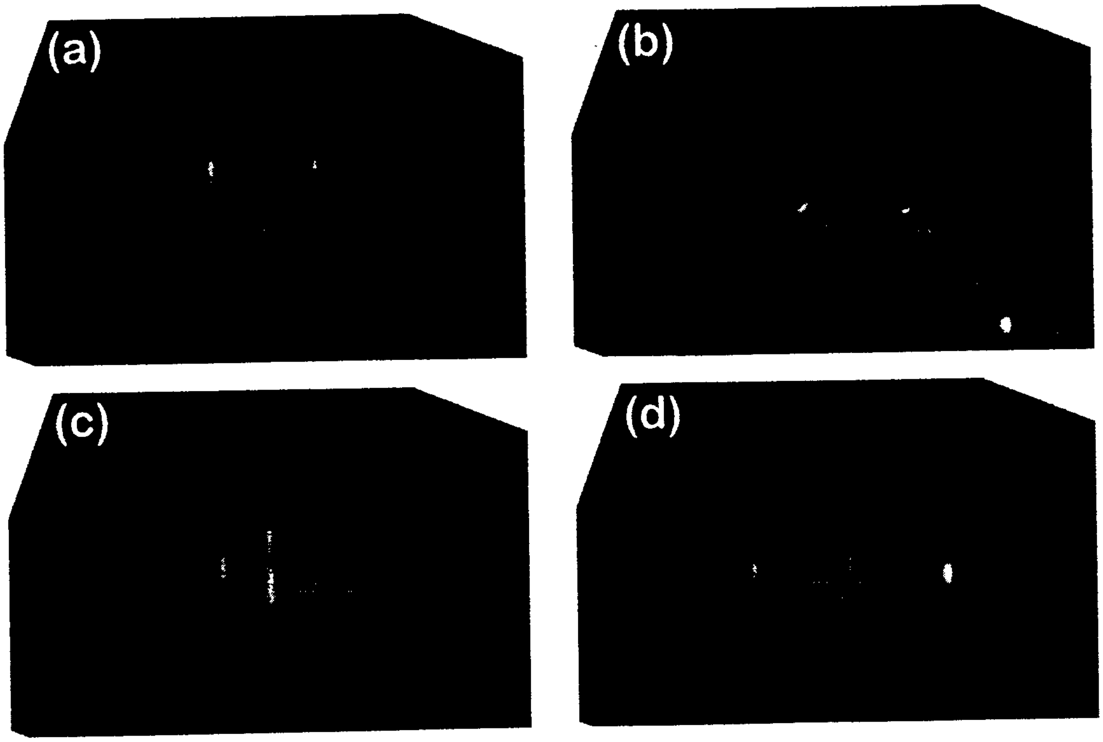


Fig. 3. RHEED patterns of a Si(1 0 0) : As 1 × 2 surface at (a) $[0\,\bar{1}\,\bar{1}]$ and (b) $[0\,\bar{1}\,\bar{1}]$ azimuthal directions. and of a BeTe 1 × 2 surface at (c) $[0\,\bar{1}\,\bar{1}]$ and (d) $[0\,\bar{1}\,\bar{1}]$ azimuthal directions.

Secondary-ion-mass-spectroscopy (SIMS) profile analysis of these samples revealed the arsenic remained at the interface between BeTe and silicon. Therefore, our RHEED results in Figs. 2 and 3 suggest an interface sequence of –Si–Si–As–B–Te–, for the BeTe layers grown on the Si(1 0 0) : As surfaces. According to an electron-accounting argument, all atoms in this interface structure are fully coordinated, therefore no interfacial dipoles are present and the structure is energetically stable [4].

When compared with the growth of ZnS, ZnSe, and GaAs on silicon surfaces, BeTe exhibited unique growth behavior of easily forming smooth surfaces. During the ZnS and ZnSe growth on Si(1 0 0) surfaces, even with the arsenic passivation layers, it was difficult to produce smooth enough surfaces for the observation of surface reconstructions by using RHEED [2, 22]. In the case of GaAs growth on silicon substrates, according to our experience, three-dimensional growth continues for several hundred to even a thousand angstroms before the surface becomes smooth. BeTe has about the same lattice mismatch that GaAs has on silicon substrates. However, for the BeTe growth on Si(1 0 0) : As surfaces, after only about three monolayers of growth, the epitaxial surface became smooth enough to observe the streaky RHEED patterns of a reconstructed surface. Even on a bare Si(1 0 0) surface, as discussed earlier, the BeTe epitaxial film became smooth within about 100–150 Å of growth.

3.3. TEM results

Fig. 4 shows a cross-sectional micrograph of a BeTe layer grown on a bare vicinal Si(1 0 0) substrate. In the BeTe/Si-substrate interface region, we observed relatively high densities of misfit dislocations "M" and stacking faults "S". The misfit dislocations were expected because of the large lattice mismatch of 3.6% between BeTe and Si. As for the stacking faults, most of them were annihilated within 150 Å. This thickness relates closely to our RHEED observation that island growth prevailed during the initial growth until about 100–150 Å. Most likely the stacking faults were formed at the peripheries of the islands when the islands collided with each other. Very few stacking faults extended to the upper surface of the 1300 Å thick BeTe layer. We also note that the crystal defect structure of the BeTe layer, shown in Fig. 4, is significantly different from that of ZnS and ZnSe layers grown on the same type of bare vicinal Si(1 0 0) surfaces [3, 4]. In both ZnS and ZnSe, the crystal defects were mostly microtwins lying exclusively in one of the two {1 1 1} planes that are parallel with the surface-step edges. Although we observed the formation of microtwins in the BeTe layer, in high-resolution micrographs which are not shown here, we did not observe any anisotropy of the stacking faults and microtwins. This suggests a different nucleation mechanism occurred for the stacking fault formation. More importantly, the stacking-fault density in the upper portion of the BeTe layer, shown in Fig. 4, is about 2 to 3 orders of magnitude smaller than those in either ZnS or ZnSe layers of similar thickness grown on the same type of vicinal Si(1 0 0) surfaces.

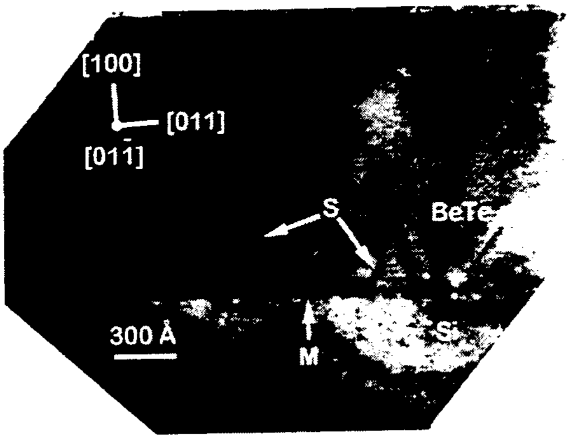


Fig. 4. Cross-sectional TEM micrograph of a BeTe layer grown on a bare vicinal Si(1 0 0) 2 × 1 surface. In the micrograph, arrows "S" point to stacking faults and arrows "M" point to misfit dislocations.

Fig. 5 shows a cross-sectional TEM micrograph of a BeTe layer grown on a Si(1 0 0) : As 2 × 1 surface. This layer showed even more improved crystal quality over the one grown on a bare Si(1 0 0) 2 × 1 surface. Very few stacking faults "S" can be observed in the region close to the epilayer/substrate interface. In this micrograph, we do not observe any stacking faults extending into or emerging from the upper portions of the layer. The stacking fault density in a BeTe layer grown on a Si(1 0 0) : As 2 × 1 surface averages about one quarter to one tenth of that in the BeTe layer grown on a bare Si(1 0 0) 2 × 1 surface. This demonstrates the effectiveness of the arsenic passivation method. Our TEM study of a BeTe layer grown on a Si(1 0 0) : As 1 × 2 surface revealed similar crystal characteristics as the one shown in Fig. 5. We also compared the stacking fault density in the BeTe layers grown on Si(1 0 0) : As surfaces with that of ZnS layers of similar thickness grown on similarly prepared Si(1 0 0) : As surfaces [3]. We found about 100 times fewer stacking faults in the BeTe layers than in the ZnS layers. From Fig. 5 we also found that all the misfit dislocations "M" are confined to the epilayer/substrate interface with no threading dislocations observed, indicating a threading dislocation density below approximately 10^7 cm^{-2}.

The TEM results in Figs. 4 and 5 establish that BeTe has a low tendency to form stacking faults. This finding is consistent with our prediction of a high stacking fault energy for the material. as shown in Fig. 1. With the optimization of the growth conditions, further improvement of the crystal quality can be expected.

The demonstration of high-quality BeTe layers grown on Si(1 0 0) : As surfaces opens up the possibility for many applications. In addition to the ones mentioned in the introduction, the material can also be used as a template or as an interlayer for the epitaxial growth of other compound semiconductors on silicon substrates. According to our estimation shown in Fig. 1, BeSe and BeS also have high stacking fault energies. Therefore, it is reasonable to expect that high crystal-quality layers of the alloys of these Be-VI compounds can be grown epitaxially on silicon substrates. These alloy layers could then form the epitaxial templates of a wide range of lattice constants.

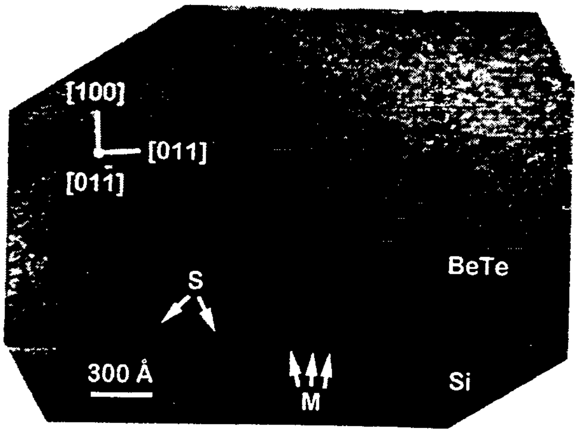


Fig. 5. Cross-sectional TEM micrograph of a BeTe layer grown on a Si(1 0 0) : As 2 × 1 surface. In the micrograph, arrows "S" point to stacking faults and arrows "M" point to misfit dislocations.

4. Conclusions

BeTe was shown to form high-quality epitaxial layers of low stacking fault densities on Si(1 0 0) substrates. On arsenic passivated Si(1 0 0) surfaces, a two-dimensional growth mode prevailed in the early growth stages in spite of a large lattice mismatch between BeTe and the silicon substrate. This study, therefore, provides a basis for the study of material fabrication and device development involving epitaxially grown Be-VI compound semiconductors.

Acknowledgements

This work was supported in part by Texas High Education Coordinating Board ARP-999903219, NSF ECS-9306293, and AFOSR/ARPA AF9620-96-C-0001. We are grateful to Robert Klima for his technical support. We acknowledge helpful discussions with Drs. R.D. Bringans of Xerox Corp. We acknowledge the Electron Microscopy Center of Texas A&M University for providing TEM facilities.

References

[1] R. Houdré and H. Morkoç, CRC Critical Rev. Solid State Mater. Sci. 16 (1990) 91.
[2] X. Zhou, S. Jiang, F. Li, G.F. Spencer, R.T. Bate and W.P. Kirk, in: Proc. 1995 IEEE Cornell Conf. on Adv. Concepts in High Speed Semi. Dev. and Circuits (1995) p. 498.

[3] L.T. Romano, R.D. Bringans, X. Zhou and W.P. Kirk, Phys. Rev. B 52 (1995) 11201.
[4] R.D. Bringans, D.K. Biegelsen, L.-E. Swartz, F.A. Ponce and J.C. Tramontana, Phys. Rev. B 45 (1992) 13400.
[5] X. Zhou and W.P. Kirk, Mater. Res. Soc. Symp. Proc. 318 (1994) 207.
[6] X. Zhou, F. Li, G.F. Spencer and W.P. Kirk, Mater. Res. Soc. Symp. Proc. 380 (1995) 3.
[7] W. Zachariasen, Z. Phys. Chem. (Leipzig) 119 (1926) 201; 124 (1926) 277, 440.
[8] W.M. Yim, J.P. Dismukes, E.J. Stofko and R.J. Paff, J. Phys. Chem. Solids 33 (1972) 501.
[9] R.L. Sarkar and S. Chatterjee, J. Phys. C 10 (1977) 57.
[10] J.C Phillips, Bonds and Bands in Semiconductors (Academic, New York. 1973).
[11] L. Takeuchi, K. Suzuki and K. Maeda, Phil. Mag. A 50 (1984) 171.
[12] J.C. Phillips and J.A. Van Vechten, Phys. Rev. Lett. 23 (1969) 1115.
[13] A. Waag. F. Fischer, H.J. Lugauer, Th. Litz, J. Laubender and G. Landwehr, Int. Symp. on Blue Light Emitting Diodes and Lasers, Chiba (1996).
[14] S. Guha. J.M. DePuydt, G.E. Hfler. M.A. Haase, B.J. Wu and H. Cheng, Appl. Phys. Lett. 63 (1993) 3023.
[15] S. Guha. J.M. DePuydt, M.A. Haase, J. Qiu and H. Cheng, Appl. Phys. Lett. 63 (1993) 3107.
[16] H. Morkoç. S. Strite, G.B. Gao, M.B. Lin, B. Sverdlov and M Burns. J. Appl. Phys. 76 (1994) 1363.
[17] C. Vèriè. in: Proc. Int. Conf. Semi. Heteroepitaxy, Montpellier. France (1995).
[18] R.C. Henderson, J. Electrochem. Soc. 119 (1972) 772.
[19] R.D. Bringans, D.K. Biegelsen and L.-E. Swartz, Phys. Rev. B 44 (1991) 3054.
[20] W.A. Harrison, E.A. Kraut, J.R. Waldrop and R.W. Grant, Phys. Rev. B 18 (1978) 4402.
[21] S. Higuchi and Y. Nakamishi, J. Appl. Phys. 71 (1992) 4277.
[22] R.D. Bringans, private conversation.